

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 06:16:23 ; Search time 21 seconds
(without alignments)
1190.943 Million cell updates/sec

Title: US-09-840-243C-11

Perfect score: 1341

Sequence: 1 MELTQPAEDLIQTQTPASE.....VIENHILKLFQSNLVPADPE 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	219.5	16.4	426	2 AE2149	hypothetical prote
2	214	16.0	2039	2 T15347	ankyrin-related un
3	207.5	15.5	1031	2 T43458	hypothetical prote
4	202.5	15.1	3924	2 S37431	ankyrin 2, neurona
5	193	14.4	4377	2 A55575	ankyrin 3, long sp
6	189	14.1	1001	2 S30385	G9a protein - huma
7	188.5	14.1	1765	2 T42714	ankyrin 3, splice
8	188.5	14.1	1940	2 T42715	ankyrin 3, splice
9	188.5	14.1	1943	2 T42713	ankyrin 3, splice
10	188.5	14.1	1961	2 T42716	ankyrin 3, splice
11	184	13.7	1881	1 SHUK	ankyrin 1, erythro
12	183	13.6	1848	2 S37771	ankyrin, erythrocy
13	183	13.6	1862	2 I49502	ankyrin - mouse
14	182	13.6	1856	2 B35049	ankyrin 1, erythro
15	182	13.6	1880	2 A35049	ankyrin 1, erythro
16	181.5	13.5	1549	2 T13940	ankyrin - fruit fl
17	176.5	13.2	2584	2 T24158	hypothetical prote
18	176.5	13.2	2606	2 T24157	hypothetical prote
19	175	13.0	495	2 T27995	probable potassium
20	173.5	12.9	888	2 D84650	ankyrin-repeat pro
21	173	12.9	333	2 JC7113	hypothetical prote
22	172.5	12.9	1188	2 T19552	2-SA-dependent RNA
23	171.5	12.8	741	2 A45771	hypothetical prote
24	171	12.8	1398	2 T21884	inv protein - mouse
25	169.5	12.6	1062	2 T14151	ankyrin repeat pro
26	167	12.5	211	2 T18174	ankyrin repeat pro
27	166.5	12.4	1062	2 T30255	inversin - mouse
28	166	12.4	658	2 S68418	protein phosphatas
29	165.5	12.3	451	1 S68455	serine/threonine-s

ALIGNMENTS

RESULT 1

AE2149

hypothetical protein all2748 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AE2149

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, T.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ar

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AE2149

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-426 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA074447.1; PID:gl7131841; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all2748

Query Match	16.4%	Score 219.5	DB 2	Length 426
Best Local Similarity	30.7%	Pred. No. 4.4e-10		
Matches	62	Conservative	34	Mismatches 71; Indels 35; Gaps 4
QY	62	QAGSSLKHSTTL	-----	TNRQGNVSVLPATLDSLSIHQLAAQGEIDQLK 107
Db	234	QDGESALHATVEGVVDVQVLLNQGANQTKNKLGDTPLLVAALQGH	-----	DQIV 285
QY	108	EHLRK-----	GDNLVKNKPDERGFTPLIWASAFGEIETVRPFLLEWGADPHILAKERSAL 161	
Db	286	ETLLKYGANVHGDNL-----	GETPLTAAASQGHATATVRIILDDYGANANIPASDGKTAL 338	
QY	162	SLATGGYTDIVGLLERVDVINYDNGGTPLLVAVRGVHKVCVEALLARGADLTTEAD 221		
Db	339	IKATERNHPGVIQLLAGANVYQDSVGATALLWAASGGYKVVQVILLEGADTNLKNR 398		
QY	222	SGYTPMDLAVALGYRKVQOVIE 243		
Db	399	GGYALMIAEFGFRSIVQILK 420		

RESULT 2

T15347

ankyrin-related unc-44 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 02-Aug-2002

C:Accession: T15347; T15346; T15344; T15345; A57282; B57282; C57282

R:Gattung, S.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid B0350.

A:Reference number: Z16332

A;Cross-references: GB:M37123; NID:g178647; PIDN:AAA62828.1; PID:g178648
P;Chan, W.; Kordeli, E.; Bennett, V.
J. Cell Biol. 123, 1463-1473, 1993
A;Title: 440-kD ankryrinB: structure of the major developmentally regulated domain and se
A;Reference number: A49462; MUID:94075409; PMID:8253844
A;Accession: A49462
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-3924 <RES>
A;Cross-references: EMBL:Z26634; NID:g406287; PIDN:CAA81387.1; PID:g406288
C;Genetics:
A;Gene: GDB:ANK2
A;Cross-references: GDB:127607; OMIM:106410
A;Map position: 4q25-q27
C;Superfamily: ankryrin; ankryrin repeat homology
C;Keywords: alternative splicing
F;2-3924/Product: ankryrin 2, long form #status predicted <MAT>
F;2-1443,3585-3924/Product: ankryrin 2, short form #status predicted <MA2>
F;63-95/Domain: ankryrin repeat homology <AN01>
F;96-128/Domain: ankryrin repeat homology <AN02>
F;129-161/Domain: ankryrin repeat homology <AN03>
F;162-190/Domain: ankryrin repeat homology <AN04>
F;191-223/Domain: ankryrin repeat homology <AN05>
F;232-264/Domain: ankryrin repeat homology <AN06>
F;265-297/Domain: ankryrin repeat homology <AN07>
F;298-330/Domain: ankryrin repeat homology <AN08>
F;331-363/Domain: ankryrin repeat homology <AN09>
F;364-396/Domain: ankryrin repeat homology <AN10>
F;397-429/Domain: ankryrin repeat homology <AN11>
F;430-462/Domain: ankryrin repeat homology <AN12>
F;463-495/Domain: ankryrin repeat homology <AN13>
F;496-528/Domain: ankryrin repeat homology <AN14>
F;529-561/Domain: ankryrin repeat homology <AN15>
F;562-594/Domain: ankryrin repeat homology <AN16>
F;595-627/Domain: ankryrin repeat homology <AN17>
F;628-660/Domain: ankryrin repeat homology <AN18>
F;661-693/Domain: ankryrin repeat homology <AN19>
F;694-726/Domain: ankryrin repeat homology <AN20>
F;727-759/Domain: ankryrin repeat homology <AN21>
F;760-792/Domain: ankryrin repeat homology <AN22>
F;793-825/Domain: ankryrin repeat homology <AN23>

Query Match 15.1%; Score 202.5; DB 2; Length 3924;
Best Local Similarity 29.3%; Pred. No. 2.5e-07;
Matches 56; Conservative 43; Mismatches 81; Indels 11; Gaps 4;

QY 55 DASVSPQAGSLKSTTTLNQRGNEVSALPATLDSLSIHQAQGLDQLKEHLRKGK 114
DB 5 DAAQKS-DSGEXFNGSSQRRKPKKSDSNA-----SFLRAARAGNLDKVVHYLKGI 55

QY 115 NLVKNKPDGRGFTPLIWASAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVG 174
DB 56 D-INTCNQNGLNALHLAAKEGHVGLVQELLGRGSSVDSATKKGNTALHIASLAGQAEVVK 114

QY 175 LLERDVIDINVDWNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLVALVG 234
DB 115 VLVKEGANINAGSQNGFTPLYNAAQENHIDVVKYLLNGANOSTATEDGFTPLAVALQGG 174

QY 235 YRK-VQOVV 244
DB 175 HNOQVAAILLEN 185

RESULT 5
A55575
ankryrin 3, long splice form - human
N;Alternate names: ankryrin G
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change 20-Sep-1999
A;Accession: A55575
R;Kordeli, E.; Lambert, S.; Bennett, V.
J. Biol. Chem. 270, 2352-2359, 1995
A;Title: Ankryrin-G. A new ankryrin gene with neural-specific isoforms localized at the ax

A;Reference number: A55575; MUID:95138209; PMID:7836469
A;Accession: A55575
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-4377 <OR>
A;Cross-references: GB:UL3616; NID:g608024; PIDN:AAA64834.1; PID:g608025
C;Genetics:
A;Gene: GDB:ANK3
A;Cross-references: GDB:424503; OMIM:600465
A;Map position: 10q21-10q21
C;Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homolog
C;Keywords: alternative splicing; peripheral membrane protein
F;73-105/Domain: ankryrin repeat homology <AN01>
F;106-138/Domain: ankryrin repeat homology <AN02>
F;139-171/Domain: ankryrin repeat homology <AN03>
F;172-200/Domain: ankryrin repeat homology <AN04>
F;201-233/Domain: ankryrin repeat homology <AN05>
F;234-266/Domain: ankryrin repeat homology <AN06>
F;267-299/Domain: ankryrin repeat homology <AN07>
F;300-332/Domain: ankryrin repeat homology <AN08>
F;333-365/Domain: ankryrin repeat homology <AN09>
F;366-398/Domain: ankryrin repeat homology <AN10>
F;399-431/Domain: ankryrin repeat homology <AN11>
F;432-464/Domain: ankryrin repeat homology <AN12>
F;465-497/Domain: ankryrin repeat homology <AN13>
F;498-530/Domain: ankryrin repeat homology <AN14>
F;531-563/Domain: ankryrin repeat homology <AN15>
F;564-596/Domain: ankryrin repeat homology <AN16>
F;597-629/Domain: ankryrin repeat homology <AN17>
F;630-662/Domain: ankryrin repeat homology <AN18>
F;663-695/Domain: ankryrin repeat homology <AN19>
F;696-728/Domain: ankryrin repeat homology <AN20>
F;729-761/Domain: ankryrin repeat homology <AN21>
F;762-794/Domain: ankryrin repeat homology <AN22>
F;795-827/Domain: ankryrin repeat homology <AN23>

Query Match 14.4%; Score 193; DB 2; Length 4377;
Best Local Similarity 28.1%; Pred. No. 1.7e-06;
Matches 50; Conservative 43; Mismatches 75; Indels 10; Gaps 3;

QY 68 KHSSTLTNRQNEVSALPATLDSLSIHQAQGLDQLKEHLRKGKDLVKNKPDGRGFTP 127
DB 27 KHRKSRDRKKKSDANA-----SYLRAARAGHLEKALDYIKNGVD-INTCNQNGLNA 77

QY 128 LIWASAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLERDVIDINVD 187
DB 78 LHLASKEGHVEVVSSELLQREANVDATKKGNTALHIASLAGQAEVVKVLTNGANVNAQS 137

QY 188 WNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLVALVG-RKQOVV 244
DB 138 QNGFTPLYNAAQENHLEVVVKFLLDNGASQSLATEDGFTPLAVALQGGHDDVSLLEN 195

RESULT 6
S30385
Gaa protein - human
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
A;Accession: S30385
R;Milner, C.M.; Campbell, R.D.
Biochem. J. 290, 811-818, 1993
A;Title: The Gaa gene in the human major histocompatibility complex encodes a novel pr
A;Reference number: S30385; MUID:93207535; PMID:8457211
A;Accession: S30385
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1001 <MIL>
A;Cross-references: EMBL:X69838; NID:g287864; PIDN:CAA49491.1; PID:g287865
C;Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homolog
F;641-673/Domain: ankryrin repeat homology <AN>

Query Match 14.1%; Score 189; DB 2; Length 1001;
Best Local Similarity 35.3%; Pred. No. 4.5e-07;

QY 134 FGIEIVRFLLEWGADPHILAKERSALSASTGGYTDIVG----- 174
Db 121 ENHLEVVKFLLENGANQNATEDGFTPLAVALQQGHENVAHLINYGTKGVRLPALHIA 180
QY 175 -----LLLRVDVINIYDNGG 191
Db 181 ARNDTRTAALLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLLNRGASVNFPTQNGI 240
QY 192 TPLLYAVRGNHVKCVKVEALLARGADLTTEADSGYTPMDLVALGYRKVQVQVIENH 245
Db 241 TPLHIASRRGNVIMVRLLDGRGAQIETRTKDELTPHCAARNGHVRISILLDH 294

RESULT 12
S37771
ankyrin, erythrocyte - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-May-1994 #sequence_revision 03-Aug-1995 #text_change 13-Aug-1999
C:Accession: S37771
R:Birkenmeier, C.S.; White, R.A.; Peters, L.L.; Hall, E.J.; Lux, S.E.; Barker, J.E.
J. Biol. Chem. 268, 9533-9540, 1993
A:Title: Complex patterns of sequence variation and multiple 5' and 3' ends are found in
A:Reference number: S37771; MUID:93252825; PMID:8486643
A:Accession: S37771
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1848 <BIR>
A:Cross-references: EMBL:X69063; NID:g311816; PIDN:CAA48801.1; PID:g311817
C:Keywords: alternative splicing
F:48-80/Domain: ankyrin repeat homology <AN01>
F:81-113/Domain: ankyrin repeat homology <AN02>
F:114-146/Domain: ankyrin repeat homology <AN03>
F:147-175/Domain: ankyrin repeat homology <AN04>
F:176-208/Domain: ankyrin repeat homology <AN05>
F:209-241/Domain: ankyrin repeat homology <AN06>
F:242-274/Domain: ankyrin repeat homology <AN07>
F:275-307/Domain: ankyrin repeat homology <AN08>
F:308-340/Domain: ankyrin repeat homology <AN09>
F:341-373/Domain: ankyrin repeat homology <AN10>
F:374-406/Domain: ankyrin repeat homology <AN11>
F:407-439/Domain: ankyrin repeat homology <AN12>
F:440-472/Domain: ankyrin repeat homology <AN13>
F:473-505/Domain: ankyrin repeat homology <AN14>
F:506-538/Domain: ankyrin repeat homology <AN15>
F:539-571/Domain: ankyrin repeat homology <AN16>
F:572-604/Domain: ankyrin repeat homology <AN17>
F:605-637/Domain: ankyrin repeat homology <AN18>
F:638-670/Domain: ankyrin repeat homology <AN19>
F:671-703/Domain: ankyrin repeat homology <AN20>
F:704-736/Domain: ankyrin repeat homology <AN21>
F:737-769/Domain: ankyrin repeat homology <AN22>
F:770-802/Domain: ankyrin repeat homology <AN23>

Query Match 13.6%; Score 183; DB 2; Length 1848;
Best Local Similarity 23.1%; Pred. No. 3.3e-06;
Matches 54; Conservative 36; Mismatches 70; Indels 74; Gaps 3;

QY 74 TNRQGNVSALPATLDSLSIHLAQOGLDQKEHLRKGDNLVNKPDERGFTPLIWA 133
Db 77 TTTKGN-----TALHIALAGQDEVVRELNYGAN-VNAOSQKGFPLYMAAQ 124
QY 134 FGIEIVRFLLEWGADPHILAKERSALSASTGGYTDIVG----- 174
Db 125 ENHLEVVKFLLENGANQNATEDGFTPLAVALQQGHENVAHLINYGTKGVRLPALHIA 184
QY 175 -----LLLRVDVINIYDNGG 191
Db 185 ARNDTRTAALLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLLNRGASVNFPTQNGI 244
QY 192 TPLLYAVRGNHVKCVKVEALLARGADLTTEADSGYTPMDLVALGYRKVQVQVIENH 245

Db 245 TPLHIASRRGNVIMVRLLDGRGAQIETRTKDELTPHCAARNGHVRISILLDH 298

RESULT 13
I49502
ankyrin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49502
R:White, R.A.; Birkenmeier, C.S.; Peters, L.L.; Barker, J.E.; Lux, S.E.
Mamm. Genome 3, 281-285, 1992
A:Title: Murine erythrocyte ankyrin cDNA: Highly conserved regions of the regulatory do
A:Reference number: I49502; MUID:92345717; PMID:1386265
A:Accession: I49502
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1862 <RES>
A:Cross-references: GB:M84756; NID:g191939; PIDN:AAA37236.1; PID:g191940
C:Genetics:
A:Gene: Ank-1
C:Superfamily: ankyrin; ankyrin repeat homology
C:Keywords: alternative splicing
F:40-72/Domain: ankyrin repeat homology <AN01>
F:73-105/Domain: ankyrin repeat homology <AN02>
F:106-138/Domain: ankyrin repeat homology <AN03>
F:139-167/Domain: ankyrin repeat homology <AN04>
F:168-200/Domain: ankyrin repeat homology <AN05>
F:201-233/Domain: ankyrin repeat homology <AN06>
F:234-266/Domain: ankyrin repeat homology <AN07>
F:267-299/Domain: ankyrin repeat homology <AN08>
F:300-332/Domain: ankyrin repeat homology <AN09>
F:333-365/Domain: ankyrin repeat homology <AN10>
F:366-398/Domain: ankyrin repeat homology <AN11>
F:399-431/Domain: ankyrin repeat homology <AN12>
F:432-464/Domain: ankyrin repeat homology <AN13>
F:465-497/Domain: ankyrin repeat homology <AN14>
F:498-530/Domain: ankyrin repeat homology <AN15>
F:531-563/Domain: ankyrin repeat homology <AN16>
F:564-596/Domain: ankyrin repeat homology <AN17>
F:597-629/Domain: ankyrin repeat homology <AN18>
F:630-662/Domain: ankyrin repeat homology <AN19>
F:663-695/Domain: ankyrin repeat homology <AN20>
F:696-728/Domain: ankyrin repeat homology <AN21>
F:729-761/Domain: ankyrin repeat homology <AN22>
F:762-794/Domain: ankyrin repeat homology <AN23>

Query Match 13.6%; Score 183; DB 2; Length 1862;
Best Local Similarity 23.1%; Pred. No. 3.3e-06;
Matches 54; Conservative 36; Mismatches 70; Indels 74; Gaps 3;

QY 74 TNRQGNVSALPATLDSLSIHLAQOGLDQKEHLRKGDNLVNKPDERGFTPLIWA 133
Db 69 TTTKGN-----TALHIALAGQDEVVRELNYGAN-VNAOSQKGFPLYMAAQ 116
QY 134 FGIEIVRFLLEWGADPHILAKERSALSASTGGYTDIVG----- 174
Db 117 ENHLEVVKFLLENGANQNATEDGFTPLAVALQQGHENVAHLINYGTKGVRLPALHIA 176
QY 175 -----LLLRVDVINIYDNGG 191
Db 177 ARNDTRTAALLQNDPNPDVLSKTGFTPLHIAAHYENLNVAQLLNRGASVNFPTQNGI 236
QY 192 TPLLYAVRGNHVKCVKVEALLARGADLTTEADSGYTPMDLVALGYRKVQVQVIENH 245
Db 237 TPLHIASRRGNVIMVRLLDGRGAQIETRTKDELTPHCAARNGHVRISILLDH 290

RESULT 14
B35049
ankyrin 1, erythrocyte splice form 3 - human
N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R
N;Contains: ankyrin 2.2, erythrocyte
C:Species: Homo sapiens (man)

C;Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 10-Jul-1998
C;Accession: B35049

R;Lambert, S.; Yu, H.; Prchal, J.T.; Lawler, J.; Ruff, P.; Speicher, D.; Cheung, M.C.; K
Proc. Natl. Acad. Sci. U.S.A. 87, 1730-1734, 1990

A;Title: cDNA sequence for human erythrocyte ankyrin.

A;Reference number: A35049; MUID:90175370; PMID:1689849

A;Accession: B35049

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1856 <LAM>

C;Genetics:

A;Gene: GDB:ANK1; ANK

A;Cross-references: GDB:118737; OMIM:182900

A;Map position: 8p11.2-8p11.2

C;Superfamily: ankyrin; ankyrin repeat homology

C;Keywords: alternative splicing

F;2-1856/Product: ankyrin 1, erythrocyte form 3 #status predicted <MAT>
F;2-1513,1676-1856/Product: ankyrin 2.2, erythrocyte #status predicted <MA2>

F;44-76/Domain: ankyrin repeat homology <AN01>

F;110-142/Domain: ankyrin repeat homology <AN03>

F;143-171/Domain: ankyrin repeat homology <AN04>

F;172-204/Domain: ankyrin repeat homology <AN05>

F;205-237/Domain: ankyrin repeat homology <AN06>

F;238-270/Domain: ankyrin repeat homology <AN07>

F;271-303/Domain: ankyrin repeat homology <AN08>

F;304-336/Domain: ankyrin repeat homology <AN09>

F;337-369/Domain: ankyrin repeat homology <AN10>

F;370-402/Domain: ankyrin repeat homology <AN11>

F;403-435/Domain: ankyrin repeat homology <AN12>

F;436-468/Domain: ankyrin repeat homology <AN13>

F;469-501/Domain: ankyrin repeat homology <AN14>

F;502-534/Domain: ankyrin repeat homology <AN15>

F;535-567/Domain: ankyrin repeat homology <AN16>

F;568-600/Domain: ankyrin repeat homology <AN17>

F;601-633/Domain: ankyrin repeat homology <AN18>

F;634-666/Domain: ankyrin repeat homology <AN19>

F;667-699/Domain: ankyrin repeat homology <AN20>

F;700-732/Domain: ankyrin repeat homology <AN21>

F;733-765/Domain: ankyrin repeat homology <AN22>

F;766-798/Domain: ankyrin repeat homology <AN23>

Query Match 13.6%; Score 182; DB 2; Length 1856;

Best Local Similarity 23.1%; Pred. No. 4e-06;

Matches 54; Conservative 37; Mismatches 69; Indels 74; Gaps 3;

QY 74 TNQRGNEVSALPATLDSLSIHQLAAGELDLQKEHLRKGDNLVKNKPDREGFTPLIWASA 133

Db 73 TTTKKGK-----TALHTAALAGQDEVVRELVNYGAN-VNAQSKGFTPLYMAAQ 120

QY 134 FGIEIVRFLLEWGADPHILAKERESALSASTGGYTDIVG----- 174

Db 121 ENHLEVVKFLLENGANQVATEDGFTPLAVALQGHENVVAHLINYGTKGKVLPAHIA 180

QY 175 -----LLLRDVIDINIDWNGG 191

Db 181 ARNDTTRTAALLQNDPNPDVLSKTGFTPLHIAAHENLVNAQLLNRGSSVNFPTQNGI 240

QY 192 TPLLYAVRGNHVKVCVALLARGADLTTEADSGYTPMDLAVALGYRKVQVVIENH 245

Db 241 TPLHIAARRGNVIMVRLLDGRGAQIETKTKDELTPHCAARNGHVRISILLDH 294

RESULT 15

A35049

ankyrin 1, erythrocyte splice form 2 - human

N;Alternate names: ankyrin 2.1, erythrocyte; ankyrin-R

N;Contains: ankyrin 2.2, erythrocyte

C;Species: Homo sapiens (man)

C;Date: 27-Jul-1990 #sequence_revision 01-Oct-1992 #text_change 04-Sep-1998

C;Accession: A35049

R;Lambert, S.; Yu, H.; Prchal, J.T.; Lawler, J.; Ruff, P.; Speicher, D.; Cheung, M.C.; K

Proc. Natl. Acad. Sci. U.S.A. 87, 1730-1734, 1990

A;Title: cDNA sequence for human erythrocyte ankyrin.

A;Reference number: A35049; MUID:90175370; PMID:1689849

A;Accession: A35049

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1880 <LAM>

A;Cross-references: GB:M28880

C;Genetics:

A;Gene: GDB:ANK1; ANK

A;Cross-references: GDB:118737; OMIM:182900

A;Map position: 8p11.2-8p11.2

C;Superfamily: ankyrin; ankyrin repeat homology

C;Keywords: alternative splicing; cytoskeleton

F;2-1880/Product: ankyrin 1, erythrocyte form 2 #status predicted <MAT>

F;2-1513,1676-1880/Product: ankyrin 2.2, erythrocyte #status predicted <MA2>

F;44-76/Domain: ankyrin repeat homology <AN01>

F;77-109/Domain: ankyrin repeat homology <AN02>

F;110-142/Domain: ankyrin repeat homology <AN03>

F;143-171/Domain: ankyrin repeat homology <AN04>

F;172-204/Domain: ankyrin repeat homology <AN05>

F;205-237/Domain: ankyrin repeat homology <AN06>

F;238-270/Domain: ankyrin repeat homology <AN07>

F;271-303/Domain: ankyrin repeat homology <AN08>

F;304-336/Domain: ankyrin repeat homology <AN09>

F;337-369/Domain: ankyrin repeat homology <AN10>

F;370-402/Domain: ankyrin repeat homology <AN11>

F;403-435/Domain: ankyrin repeat homology <AN12>

F;436-468/Domain: ankyrin repeat homology <AN13>

F;469-501/Domain: ankyrin repeat homology <AN14>

F;502-534/Domain: ankyrin repeat homology <AN15>

F;535-567/Domain: ankyrin repeat homology <AN16>

F;568-600/Domain: ankyrin repeat homology <AN17>

F;601-633/Domain: ankyrin repeat homology <AN18>

F;634-666/Domain: ankyrin repeat homology <AN19>

F;667-699/Domain: ankyrin repeat homology <AN20>

F;700-732/Domain: ankyrin repeat homology <AN21>

F;733-765/Domain: ankyrin repeat homology <AN22>

F;766-798/Domain: ankyrin repeat homology <AN23>

Query Match 13.6%; Score 182; DB 2; Length 1880;

Best Local Similarity 23.1%; Pred. No. 4.1e-06;

Matches 54; Conservative 37; Mismatches 69; Indels 74; Gaps 3;

QY 74 TNQRGNEVSALPATLDSLSIHQLAAGELDLQKEHLRKGDNLVKNKPDREGFTPLIWASA 133

Db 73 TTTKKGK-----TALHTAALAGQDEVVRELVNYGAN-VNAQSKGFTPLYMAAQ 120

QY 134 FGIEIVRFLLEWGADPHILAKERESALSASTGGYTDIVG----- 174

Db 121 ENHLEVVKFLLENGANQVATEDGFTPLAVALQGHENVVAHLINYGTKGKVLPAHIA 180

QY 175 -----LLLRDVIDINIDWNGG 191

Db 181 ARNDTTRTAALLQNDPNPDVLSKTGFTPLHIAAHENLVNAQLLNRGSSVNFPTQNGI 240

QY 192 TPLLYAVRGNHVKVCVALLARGADLTTEADSGYTPMDLAVALGYRKVQVVIENH 245

Db 241 TPLHIAARRGNVIMVRLLDGRGAQIETKTKDELTPHCAARNGHVRISILLDH 294

Search completed: April 29, 2004, 06:27:18

Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run On: April 29, 2004, 06:24:08 ; Search time 23 Seconds
(without alignments)
583.598 Million cell updates/sec

Title: US-09-840-243c-11
Perfect score: 1341
Sequence: 1 MELTQPAEDLIQTQTPASE.....VIENHILKLFQSNLVPADPE 260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2.6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2.6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2.6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2.6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2.6/prodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2.6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1341	100.0	260	2	US-09-172-977-1
2	1341	100.0	260	4	US-09-404-108-1
3	1113.5	83.0	269	4	US-09-339-746A-2
4	205	15.3	348	2	US-09-031-485-28
5	205	15.3	348	2	US-08-847-429A-28
6	205	15.3	348	3	US-09-065-474-28
7	205	15.3	348	4	US-09-557-034-28
8	205	15.3	1745	2	US-09-031-485-33
9	205	15.3	1745	2	US-08-847-429A-33
10	205	15.3	1745	3	US-09-065-474-33
11	205	15.3	1745	4	US-09-557-034-33
12	202.5	15.1	1839	2	US-09-172-977-4
13	202.5	15.1	1839	4	US-09-404-108-4
14	197	14.7	843	2	US-09-172-977-3
15	197	14.7	843	4	US-09-404-108-3
16	187.5	14.0	1088	3	US-09-082-059-2
17	187	13.9	673	3	US-09-196-387-8
18	187	13.9	673	4	US-09-841-835-8
19	187	13.9	949	3	US-09-196-387-10
20	187	13.9	949	4	US-09-841-835-10
21	187	13.9	1327	3	US-09-196-387-2
22	187	13.9	1327	4	US-09-841-835-2
23	187	13.9	1327	4	US-09-972-115A-8
24	179	13.3	994	4	US-10-164-595-38
25	175	13.0	33	4	US-09-339-746A-3
26	175	13.0	435	4	US-09-533-029-56
27	172.5	12.9	1619	4	US-09-392-812A-4

28	171.5	12.8	741	2	US-08-462-481-2	Sequence 2, Appli
29	171.5	12.8	741	2	US-08-436-771-2	Sequence 2, Appli
30	171.5	12.8	741	2	US-08-434-998-2	Sequence 2, Appli
31	171.5	12.8	741	2	US-08-487-797-2	Sequence 2, Appli
32	171.5	12.8	741	2	US-08-701-005A-2	Sequence 2, Appli
33	171.5	12.8	741	2	US-08-479-895-2	Sequence 2, Appli
34	171.5	12.8	741	3	US-08-943-956A-2	Sequence 2, Appli
35	171.5	12.8	741	5	PCT-US95-02058-2	Sequence 2, Appli
36	170.5	12.7	228	4	US-09-543-681A-4781	Sequence 4781, Ap
37	169.5	12.6	741	2	US-08-436-771-4	Sequence 4, Appli
38	169.5	12.6	741	2	US-08-434-998-4	Sequence 4, Appli
39	169.5	12.6	741	2	US-08-487-797-4	Sequence 4, Appli
40	169.5	12.6	741	5	PCT-US95-02058-4	Sequence 4, Appli
41	166.5	12.4	248	4	US-09-328-352-6520	Sequence 6520, Ap
42	165.5	12.3	345	4	US-09-252-991A-17809	Sequence 17809, A
43	165.5	12.3	452	3	US-09-035-706-2	Sequence 2, Appli
44	165.5	12.3	452	3	US-08-955-841-2	Sequence 2, Appli
45	165.5	12.3	452	4	US-09-390-425-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-172-977-1
; Sequence 1, Application US/09172977
; Patent No. 5989863
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: HUMAN ANKYRIN FAMILY PROTEIN
; FILE REFERENCE: PF-0615 US
; CURRENT APPLICATION NUMBER: US/09/172,977
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1808075
US-09-172-977-1

Query Match	100.0%	Score 1341;	DB 2;	Length 260;
Best Local Similarity	100.0%	Pred. No. 1.2e-129;		
Matches 260;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MELTQPAEDLIQTQTPASELGDPEGEAAGDSGTVVLSLFPCTPEVPNPEDASVSS	60	
DB	1	MELTQPAEDLIQTQTPASELGDPEGEAAGDSGTVVLSLFPCTPEVPNPEDASVSS	60	
QY	61	POAGSLKHSTTLTRQRNEVSALPATLDSLSIHLAAQGGELDQKEHLRGDNLVKNP	120	
DB	61	POAGSLKHSTTLTRQRNEVSALPATLDSLSIHLAAQGGELDQKEHLRGDNLVKNP	120	
QY	121	DERGFTPLIWSAFGEIETVRFLEWGDPPHILAKERESALSLASTGGYTDIVGLLLERD	180	
DB	121	DERGFTPLIWSAFGEIETVRFLEWGDPPHILAKERESALSLASTGGYTDIVGLLLERD	180	
QY	181	VDINITYDMNGGTFPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVLGYRKVQO	240	
DB	181	VDINITYDMNGGTFPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVLGYRKVQO	240	
QY	241	VIENHILKLFQSNLVPADPE	260	
DB	241	VIENHILKLFQSNLVPADPE	260	

RESULT 2
US-09-404-108-1

; Sequence 1, Application US/09404108
; Patent No. 6590077
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guigler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: HUMAN ANKYRIN FAMILY PROTEIN
; FILE REFERENCE: PF-0615 US
; CURRENT APPLICATION NUMBER: US/09/404,108
; CURRENT FILING DATE: 1999-09-23
; EARLIER APPLICATION NUMBER: 09/172,977
; EARLIER FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1808075
US-09-404-108-1

Query Match 100.0%; Score 1341; DB 4; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.2e-129;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MELTQPAEDLIQTQOTPASELGDPEDGEEADGSDTVVLSLFPCTPPEPVPNPEDASVSS 60
Db 1 MELTQPAEDLIQTQOTPASELGDPEDGEEADGSDTVVLSLFPCTPPEPVPNPEDASVSS 60
QY 61 PQAGSSLKHSHTLTNRQRGNEVSALPATLDSLSIHQAAQGLDQLKEHLRKGDNLVNKP 120
Db 61 PQAGSSLKHSHTLTNRQRGNEVSALPATLDSLSIHQAAQGLDQLKEHLRKGDNLVNKP 120
QY 121 DERGFTPLIWSAFGEIETVRFLEWGADPHILAKERSALSASTGGYTDIVGLLLERD 180
Db 121 DERGFTPLIWSAFGEIETVRFLEWGADPHILAKERSALSASTGGYTDIVGLLLERD 180
QY 181 VDINIYDNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKVQ 240
Db 181 VDINIYDNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKVQ 240
QY 241 VIENHILKLFQSNLVPADPE 260
Db 241 VIENHILKLFQSNLVPADPE 260

RESULT 3
US-09-339-746A-2
; Sequence 2, Application US/09339746A
; Patent No. 6617427
; GENERAL INFORMATION:
; APPLICANT: Tschlis, Philip N
; APPLICANT: Makris, Antonios
; TITLE OF INVENTION: Nucleic Acid Molecule Encoding An Ankyrin Repeat Tvl-1 Protein An
; FILE REFERENCE: 99-40123US
; CURRENT APPLICATION NUMBER: US/09/339,746A
; CURRENT FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: US 60/090,742
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Amino acid sequence of the murine Tvl-1 protein
US-09-339-746A-2

Query Match 83.0%; Score 1113.5; DB 4; Length 269;

Best Local Similarity 81.9%; Pred. No. 3.2e-106;
Matches 221; Conservative 15; Mismatches 23; Indels 11; Gaps 2;
QY 1 MELTQPAEDLIQTQOTPASELGDPEDGEEADGSDTVVLSLFPCTPPEPVPNPEDASVSS 60
Db 1 MEPTQVAENLVNQPFPVDDLEDDTDESPENDTVVLSLFPCTPDAVNEADASASS 60
QY 61 PQAGSSLKHSHTLTNRQRGNEVSALPATLDSLSIHQAAQGLDQLKEHLRKGACPACTC 119
Db 61 LQ-GSFLKHSHTLTNRQRGNEVSALPATLDSLSIHQAAQGLDQLKEHLRKGACPACTC 119
QY 113 --GDNLVNKPDERGFTPLIWSAFGEIETVRFLEWGADPHILAKERSALSASTGGY 170
Db 120 LSGNNLINKPDERGFTPLIWSAFGEIETVRFLEWGADPHILAKERSALSASTGGY 179
QY 171 DIVGLLLERDVIINIYDNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLA 230
Db 180 DIVRLLDRDVIINIYDNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLA 239
QY 231 VALGYRKVQVIENHILKLFQSNLVPADPE 260
Db 240 VALGYRKVQVWESHILRFLQSTLGFVDPE 269

RESULT 4
US-09-031-485-28
; Sequence 28, Application US/09031485
; Patent No. 5824306
; GENERAL INFORMATION:
; APPLICANT: Tang, Liang
; APPLICANT: Biehm, E. Scott
; TITLE OF INVENTION: DIFOLARIA AND BRUGIA ANKYRIN
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carol Talkington Verser, Ph.D.
; ADDRESSEE: Heska Corporation
; STREET: 1825 Sharp Point Drive
; CITY: Fort Collins
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80525
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WordPerfect for Windows, Version 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,485
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,429
; FILING DATE: 24-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Verser, Carol Talkington
; REGISTRATION NUMBER: 37,459
; REFERENCE/DOCKET NUMBER: HW-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 970/493-7272
; TELEFAX: 970/484-9505
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-031-485-28

Query Match 15.3%; Score 205; DB 2; Length 348;
Best Local Similarity 33.9%; Pred. No. 1.2e-12;
Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;

QY 75 NRQRGNEVSALPATLDSLSIHQAAQGGELDQLKEHLRKGDNLVNKPDERGFTPLIWASAF 134
 DB 28 NSQHSNKGES-----SASFRAARAGNLDRLVLELLRSCTD-INTCNANGLNALHLASKE 80
 QY 135 GEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERDVIDINIYDNGGTPL 194
 DB 81 GHHEVVRELLKRAVDVDAATRKGNTHALHTASLAGQELIVTVLVENGANVNVQSLNGFTPL 140
 QY 195 LYAVRGNHVKCVALLARGADLTTEADSGYTPMDLVALGY-RKVQOVLEN 244
 DB 141 YMAAQENHESVRYLLAHNANQALSTEDGFTPLAVALQQGHDRVAVLLEN 191

RESULT 5

US-08-847-429A-28
 ; Sequence 28, Application US/08847429A
 ; Patent No. 5827692
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Liang
 ; APPLICANT: Blehm, E. Scot
 ; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
 ; TITLE OF INVENTION: USES THEREOF
 ; NUMBER OF SEQUENCES: 85
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Carol Talkington Verser, Ph.D.
 ; ADDRESSEE: Heska Corporation
 ; STREET: 1825 Sharp Point Drive
 ; CITY: Fort Collins
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80525

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: WordPerfect for Windows, Version 7.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/847,429A
 FILING DATE: 24-APR-1997
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Verser, Carol Talkington
 REGISTRATION NUMBER: 37,459
 REFERENCE/DOCKET NUMBER: HW-5
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 970/493-7272
 TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 348 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-847-429A-28

Query Match 15.3%; Score 205; DB 2; Length 348;
 Best Local Similarity 33.9%; Pred. No. 1.2e-12;
 Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;

QY 75 NRQRGNEVSALPATLDSLSIHQAAQGGELDQLKEHLRKGDNLVNKPDERGFTPLIWASAF 134
 DB 28 NSQHSNKGES-----SASFRAARAGNLDRLVLELLRSCTD-INTCNANGLNALHLASKE 80
 QY 135 GEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERDVIDINIYDNGGTPL 194
 DB 81 GHHEVVRELLKRAVDVDAATRKGNTHALHTASLAGQELIVTVLVENGANVNVQSLNGFTPL 140
 QY 195 LYAVRGNHVKCVALLARGADLTTEADSGYTPMDLVALGY-RKVQOVLEN 244
 DB 141 YMAAQENHESVRYLLAHNANQALSTEDGFTPLAVALQQGHDRVAVLLEN 191

RESULT 6

US-09-065-474-28
 ; Sequence 28, Application US/09065474
 ; Patent No. 6063599
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Liang
 ; APPLICANT: Blehm, E. Scot
 ; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
 ; TITLE OF INVENTION: USES THEREOF
 ; NUMBER OF SEQUENCES: 171
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Carol Talkington Verser, Ph.D.
 ; ADDRESSEE: Heska Corporation
 ; STREET: 1825 Sharp Point Drive
 ; CITY: Fort Collins
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80525

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: WordPerfect for Windows, Version 7.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/065,474
 FILING DATE: 24-APR-1998
 CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
 NAME: Verser, Carol Talkington
 REGISTRATION NUMBER: 37,459
 REFERENCE/DOCKET NUMBER: HW-5-C1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 970/493-7272
 TELEFAX: 970/484-9505
 INFORMATION FOR SEQ ID NO: 28:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 348 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-065-474-28

Query Match 15.3%; Score 205; DB 3; Length 348;
 Best Local Similarity 33.9%; Pred. No. 1.2e-12;
 Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;

QY 75 NRQRGNEVSALPATLDSLSIHQAAQGGELDQLKEHLRKGDNLVNKPDERGFTPLIWASAF 134
 DB 28 NSQHSNKGES-----SASFRAARAGNLDRLVLELLRSCTD-INTCNANGLNALHLASKE 80
 QY 135 GEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERDVIDINIYDNGGTPL 194
 DB 81 GHHEVVRELLKRAVDVDAATRKGNTHALHTASLAGQELIVTVLVENGANVNVQSLNGFTPL 140
 QY 195 LYAVRGNHVKCVALLARGADLTTEADSGYTPMDLVALGY-RKVQOVLEN 244
 DB 141 YMAAQENHESVRYLLAHNANQALSTEDGFTPLAVALQQGHDRVAVLLEN 191

RESULT 7

US-09-557-034-28
 ; Sequence 28, Application US/09557034
 ; Patent No. 6365569
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang, Liang
 ; APPLICANT: Blehm, E. Scot
 ; TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
 ; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
 ; TITLE OF INVENTION: USES THEREOF
 ; NUMBER OF SEQUENCES: 171
 ; CORRESPONDENCE ADDRESS:

```
/
/ ADDRESSER: Carol Talkington Verser, Ph.D.
/ Heskia Corporation
/ STREET: 1825 Sharp Point Drive
/ CITY: Fort Collins
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80525
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/557,034
/ FILING DATE: 21-Apr-2000
/ CLASSIFICATION: <Unknown>
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 09/065,474
/ FILING DATE: <Unknown>
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Verser, Carol Talkington
/ REGISTRATION NUMBER: 37,459
/ REFERENCE/DOCKET NUMBER: HW-5-C1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 970/493-7272
/ TELEFAX: 970/484-9505
/ INFORMATION FOR SEQ ID NO: 28:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 348 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 28:
/
/ US-09-557-034-28
/
/ Query Match 15.3%; Score 205; DB 4; Length 348;
/ Best Local Similarity 33.9%; Pred. No. 1.2e-12;
/ Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;
/
/ QY 75 NRQGNVSALPATLDSLSIHQAQGLDQKHEHLRGKGNLVNKPDERGFTPLIWASAF 134
/ Db 28 NSQHSNKGES-----SASFLLRAARAGNLDRLVLELRSGTD-INTCNANGLNALHLSAKE 80
/
/ QY 135 GEIETVRFLEWGADPHILAKERESALSLASTGGYTDIVGLLLERDVIDINIYDMNGGTPL 194
/ Db 81 GHHEVVRELLKRAKDVAATRKGNVTALHSLAGQELIVTVLVENGANVNVQSLNGFTPL 140
/
/ QY 195 LYAVRGNHVKVCEALLARGADLTTEADSGYTPMDLAVAGY-RKVOQVNIEN 244
/ Db 141 YMAAQENHESVVRVYLLAHNANQALSTEDGFTPLAVALQQGHDRVAVLLEN 191
/
/ RESULT 8
/ US-09-031-485-33
/ Sequence 33, Application US/09031485
/ Patent No. 5824306
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Liang
/ APPLICANT: Blehm, E. Scot
/ TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
/ TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
/ NUMBER OF SEQUENCES: 85
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Carol Talkington Verser, Ph.D.
/ ADDRESSEE: Heskia Corporation
/ STREET: 1825 Sharp Point Drive
/ CITY: Fort Collins
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80525
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/557,034
/ FILING DATE: 21-Apr-2000
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Verser, Carol Talkington
/ REGISTRATION NUMBER: 37,459
/ MEDIUM TYPE: Floppy disk
```

```
/
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/031,485
/ FILING DATE:
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/847,429
/ FILING DATE: 24-Apr-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Verser, Carol Talkington
/ REGISTRATION NUMBER: 37,459
/ REFERENCE/DOCKET NUMBER: HW-5
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 970/493-7272
/ TELEFAX: 970/484-9505
/ INFORMATION FOR SEQ ID NO: 33:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1745 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/
/ US-09-031-485-33
/
/ Query Match 15.3%; Score 205; DB 2; Length 1745;
/ Best Local Similarity 33.9%; Pred. No. 1.4e-11;
/ Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;
/
/ QY 75 NRQGNVSALPATLDSLSIHQAQGLDQKHEHLRGKGNLVNKPDERGFTPLIWASAF 134
/ Db 28 NSQHSNKGES-----SASFLLRAARAGNLDRLVLELRSGTD-INTCNANGLNALHLSAKE 80
/
/ QY 135 GEIETVRFLEWGADPHILAKERESALSLASTGGYTDIVGLLLERDVIDINIYDMNGGTPL 194
/ Db 81 GHHEVVRELLKRAKDVAATRKGNVTALHSLAGQELIVTVLVENGANVNVQSLNGFTPL 140
/
/ QY 195 LYAVRGNHVKVCEALLARGADLTTEADSGYTPMDLAVAGY-RKVOQVNIEN 244
/ Db 141 YMAAQENHESVVRVYLLAHNANQALSTEDGFTPLAVALQQGHDRVAVLLEN 191
/
/ RESULT 9
/ US-08-847-429A-33
/ Sequence 33, Application US/08847429A
/ Patent No. 5827692
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Liang
/ APPLICANT: Blehm, E. Scot
/ TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN
/ TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
/ NUMBER OF SEQUENCES: 85
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Carol Talkington Verser, Ph.D.
/ ADDRESSEE: Heskia Corporation
/ STREET: 1825 Sharp Point Drive
/ CITY: Fort Collins
/ STATE: Colorado
/ COUNTRY: USA
/ ZIP: 80525
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: WordPerfect for Windows, Version 7.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/847,429A
/ FILING DATE: 24-Apr-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Verser, Carol Talkington
/ REGISTRATION NUMBER: 37,459
```

REFERENCE/DOCKET NUMBER: HW-5
TELECOMMUNICATION INFORMATION:
TELEPHONE: 970/493-7272
TELEFAX: 970/484-9505
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 1745 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-847-429A-33

Query Match 15.3%; Score 205; DB 2; Length 1745;

Best Local Similarity 33.9%; Pred. No. 1.4e-11;
Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;

QY 75 NRQRNEVSALPATLDSLSIHQAAQGLDOLKEHLKRGDNLVKNKPDGRGFTPLIWASAF 134
DB 28 NSQHSNKGES-----SASFLEAARAGNLDRLVLELLRSSTD-INTCNANGLNALHLASKE 80
QY 135 GEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERDVIDINIYDWMNGTPL 194
DB 81 GHHEVVRELLKRAVDAAATRGNTALHIALSLAGQELIVTVLVENGANVNVQSLNGFTPL 140
QY 195 LYAVRGNHVKVEALLARGADLTTEADSGYTPMDLVAALGY-RKVOQVNIEN 244
DB 141 YMAAQENHESVVRVYLLAHNAQALSTEDGFTPLAVALQQGHDVRVAVLLEN 191

RESULT 10

US-09-065-474-33
Sequence 33, Application US/09065474
Patent No. 6063599

GENERAL INFORMATION:
APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scott

TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN

TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES, AND
US THEREOF

NUMBER OF SEQUENCES: 171

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: WordPerfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/065,474

FILING DATE: 24-APR-1998

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: HW-5-C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 1745 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-065-474-33

Query Match

15.3%; Score 205; DB 3; Length 1745;

Best Local Similarity 33.9%; Pred. No. 1.4e-11;
Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;

QY 75 NRQRNEVSALPATLDSLSIHQAAQGLDOLKEHLKRGDNLVKNKPDGRGFTPLIWASAF 134

DB 28 NSQHSNKGES-----SASFLEAARAGNLDRLVLELLRSSTD-INTCNANGLNALHLASKE 80

QY 135 GEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERDVIDINIYDWMNGTPL 194

DB 81 GHHEVVRELLKRAVDAAATRGNTALHIALSLAGQELIVTVLVENGANVNVQSLNGFTPL 140

QY 195 LYAVRGNHVKVEALLARGADLTTEADSGYTPMDLVAALGY-RKVOQVNIEN 244

DB 141 YMAAQENHESVVRVYLLAHNAQALSTEDGFTPLAVALQQGHDVRVAVLLEN 191

RESULT 11

US-09-557-034-33

Sequence 33, Application US/09557034

Patent No. 6365569

GENERAL INFORMATION:

APPLICANT: Tang, Liang

APPLICANT: Blehm, E. Scott

TITLE OF INVENTION: DIROFILARIA AND BRUGIA ANKYRIN

PROTEINS, NUCLEIC ACID MOLECULES, AND

US THEREOF

NUMBER OF SEQUENCES: 171

CORRESPONDENCE ADDRESS:

ADDRESSEE: Carol Talkington Verser, Ph.D.

STREET: 1825 Sharp Point Drive

CITY: Fort Collins

STATE: Colorado

COUNTRY: USA

ZIP: 80525

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: Windows 95

SOFTWARE: WordPerfect for Windows, Version 7.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,034

FILING DATE: 21-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/065,474

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Verser, Carol Talkington

REGISTRATION NUMBER: 37,459

REFERENCE/DOCKET NUMBER: HW-5-C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 970/493-7272

TELEFAX: 970/484-9505

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 1745 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 33:

US-09-557-034-33

Query Match

15.3%; Score 205; DB 4; Length 1745;

Best Local Similarity 33.9%; Pred. No. 1.4e-11;

Matches 58; Conservative 28; Mismatches 77; Indels 8; Gaps 3;

QY 75 NRQRNEVSALPATLDSLSIHQAAQGLDOLKEHLKRGDNLVKNKPDGRGFTPLIWASAF 134

DB 28 NSQHSNKGES-----SASFLEAARAGNLDRLVLELLRSSTD-INTCNANGLNALHLASKE 80

QY 135 GEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERDVIDINIYDWMNGTPL 194


```

; TITLE OF INVENTION: HUMAN ANKYRIN FAMILY PROTEIN
; FILE REFERENCE: PF-0615 US
; CURRENT APPLICATION NUMBER: US/09/404,108
; CURRENT FILING DATE: 1999-09-23
; EARLIER APPLICATION NUMBER: 09/172,977
; EARLIER FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PERL Program
; SEQ ID NO 3
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE: -
; OTHER INFORMATION: g1841966
US-09-404-108-3

```

```

Query Match      14.7%; Score 197; DB 4; Length 843;
Best Local Similarity 32.4%; Pred. No. 3e-11;
Matches 48; Conservative 35; Mismatches 63; Indels 2; Gaps 2;

QY 98 AAQGEIDQLKEHLKGDNLVKNKPDGFTPLIWSAFGEIETVRFLEWGADPHIIAKER 157
Db 4 ARAGNLDKVVVEYLKGGID-INTCNQNGLNALHLAAKEGHVGLVQELLGRGSSVDSATKKG 62

QY 158 ESALSIASTGGYTDIVGLLLERDVIDINIYDWNNGTPLLVAVRGNHVKCVVEALLARGADLT 217
Db 63 NTALHIASLAGQAQAEVVKVLVKEGANINAQSQNGFTPLYNAAQENHIDVVVKYLLLENGANOS 122

QY 218 TEADSGYTPMDLAVALGYEK-VQQVIEN 244
Db 123 TATEDGFTPLAVALQQGHQNAVAILLEN 150

```

Search completed: April 29, 2004, 06:27:53
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 06:16:22 ; Search time 60 Seconds
(without alignments)
1224.373 Million cell updates/sec

Title: US-09-840-243c-11
Perfect score: 1341
Sequence: 1 MELTQPAEDLIQTQTPASE.....VIENHLKLFQSNLVPADPE 260

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1341	100.0	260	3 AAY59539	Human ank
2	1341	100.0	260	3 AAY79411	Human MHC
3	1341	100.0	260	7 ADD27859	Human ank
4	1113.5	83.0	269	3 AAY59590	Ankyrin r
5	625	46.6	229	5 ABP41881	Human ova
6	621.5	46.3	313	4 AAG66309	Human ank
7	621.5	46.3	313	4 AAB94322	Human pro
8	621.5	46.3	314	7 ADE60905	Human pro
9	621.5	46.3	314	7 ADE60909	Human pro
10	612.5	45.7	312	7 ADD47704	Rat Prote
11	607.5	45.3	263	4 AAU20646	Human sec
12	495	36.9	152	4 ABB50161	Human tra
13	421.5	31.4	84	3 AAG01584	Human sec
14	393	29.3	119	4 AAU20665	Human sec
15	364	27.1	105	4 AAU20558	Human sec
16	299	22.3	81	5 ABB97342	Novel hum
17	274	20.4	234	4 ABB61859	Drosophil
18	257	19.2	49	4 AAM21225	Peptide #
19	257	19.2	49	4 ABB43550	Peptide #
20	257	19.2	49	5 ABG46308	Human pep
21	216.5	16.1	747	6 ADA54356	Human pro
22	216.5	16.1	1429	6 AAE36185	Human CGD
23	212	15.8	705	4 AAG75604	Human col
24	212	15.8	1184	7 ADC06846	Human kid
25	212	15.8	1762	5 AAU96841	Rat kidin

26	212	15.8	1762	7 ADC06848	Adc06848 Rat Kidin
27	212	15.8	1771	7 ADC06847	Adc06847 Human Kid
28	210.5	15.7	2443	4 ABB60521	Abb60521 Drosophil
29	207.5	15.5	166	5 ABB78585	Abb78585 3 ankyrin
30	207.5	15.5	1142	7 ADC06850	Adc06850 Kidins 22
31	207.5	15.5	1715	4 AAM39025	Aam39025 Human pol
32	207.5	15.5	1715	4 AAM38993	Aam38993 Human pol
33	207.5	15.5	1715	5 AAE25144	Aae25144 Human ARM
34	207.5	15.5	1715	5 AAE25143	Aae25143 Rat ARMS
35	207.5	15.5	1715	5 AAU96840	Aau96840 Human kid
36	207.5	15.5	1715	6 AAE32128	Aae32128 Human cyt
37	207.5	15.5	1715	7 ADC06849	Adc06849 Rat Kidin
38	207.5	15.5	1753	6 ADA09888	Ada09888 Human rec
39	206.5	15.4	1763	5 AAB080244	Abb080244 Rat Kidin
40	205.5	15.3	342	4 ABB59641	Abb59641 Drosophil
41	205	15.3	348	2 AAW76775	Aaw76775 D. immiti
42	205	15.3	348	2 AAW70607	Aaw70607 Ankyrin p
43	205	15.3	348	3 AAB11588	Aab11588 D. immiti
44	205	15.3	348	5 AAC021367	Aac021367 Ankyrin p
45	205	15.3	1745	2 AAW76776	Aaw76776 D. immiti

ALIGNMENTS

RESULT 1

AA59539
ID AAY59539 standard; protein; 260 AA.

AC AAY59539;

DT 03-APR-2000 (first entry)

DE Human ankyrin family protein, ANFP.

XX Human; ankyrin family protein; ANFP; autoimmune disorder; inflammation; atherosclerosis; inflammatory disorder; proliferative disorder; AIDS;
XX vesicle-trafficking disorder; allergy; amyloidosis; anaemia; asthma;
XX bronchitis; Crohn's disease; atopic dermatitis; diabetes mellitus;
XX irritable bowel syndrome; osteoporosis; rheumatoid arthritis; cirrhosis;
XX hepatitis; ulcerative colitis; cancer; hypercholesterolaemia; therapy;
XX diagnosis.

OS Homo sapiens.

XX US5989863-A.

XX 23-NOV-1999.

XX 14-OCT-1998; 98US-00172977.

XX 14-OCT-1998; 98US-00172977.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Corley NC, Yue H, Guegler KJ;

XX WPI; 2000-095634/08.

XX N-PSDB; AA249052.

XX Polynucleotide sequence encoding a human ankyrin family protein useful for diagnosis or treatment of autoimmune, inflammatory, proliferative and vesicle-trafficking disorders.

XX Claim 1; Fig 1; 34pp; English.

XX This sequence is the human ankyrin family protein, ANFP, of the invention. Host cells containing an expression vector containing the polynucleotide sequence can be cultured to produce ANFP, which can be used for diagnosis or treatment of autoimmune, inflammatory, proliferative and vesicle-trafficking disorders. Disorders which can be treated include acquired immune deficiency syndrome (AIDS), allergies, amyloidosis, anaemia, asthma, atherosclerosis, bronchitis, Crohn's

CC disease, atopic dermatitis, diabetes mellitus, irritable bowel syndrome,
CC myocardial or pericardial inflammation, osteoporosis, rheumatoid
CC arthritis, cirrhosis, hepatitis, ulcerative colitis, cancer and
CC hypercholesterolaemia. The polynucleotide sequences can also be used as a
CC hybridisation probe to detect ANFP-encoding polynucleotides in biological
CC samples. Purified ANFP can be used to produce antibodies or to screen
CC libraries of pharmaceutical agents to find agents that specifically bind
CC ANFP. The DNA and its antisense sequence can be used in therapeutic
CC compositions e.g. to regulate gene function. The DNA sequence can be used
CC for diagnostic purposes to detect and quantitate gene expression in
CC biopsied tissues and to indicate the absence, presence and excess
CC expression of ANFP and monitor its levels during therapeutic intervention
XX
XX
SQ Sequence 260 AA;

Query Match 100.0%; Score 1341; DB 3; Length 260;
Best Local Similarity 100.0%; Pred. No. 5.9e-128;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MELTQPAEDLIQTQTPASELGDPEDEGAEADGSDTVVLSLFPCTPEPVNPEPDASVSS 60
DB 1 MELTQPAEDLIQTQTPASELGDPEDEGAEADGSDTVVLSLFPCTPEPVNPEPDASVSS 60
QY 61 PQAGSLKHSTTLTNQRGNEVSALPATLDSLSIHQAAQGLDQKHELRKGDNLVNP 120
DB 61 PQAGSLKHSTTLTNQRGNEVSALPATLDSLSIHQAAQGLDQKHELRKGDNLVNP 120
QY 121 DERGFPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERD 180
DB 121 DERGFPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERD 180
QY 181 VDINIYDNGGTPLLYAVRGNHVKCVALLARGADLTTEADSGYTPMDLAVALGYRKVQ 240
DB 181 VDINIYDNGGTPLLYAVRGNHVKCVALLARGADLTTEADSGYTPMDLAVALGYRKVQ 240
QY 241 VIENHILKLFQSNLVPADPE 260
DB 241 VIENHILKLFQSNLVPADPE 260

RESULT 2
AA79411
ID AA79411 standard; protein; 260 AA.
AC AA79411;
XX
XX
DT 01-AUG-2000 (first entry)
XX
XX
DE Human MHC class II gene transcription factor RFXANK.
XX
XX
KW RFXANK; HsRFXANK; human; transcription factor; MHC class II;
KW chromosome 1p12; immunosuppressive; immunomodulator; antiinflammatory;
KW antidiabetic; antiarthritic; therapy; inflammation; autoimmune diseases;
KW transplant rejection; insulin dependent diabetes; multiple sclerosis;
KW lupus erythematosus; rheumatoid arthritis; immunodeficiency.
XX
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT Region 122..156
FT /note= "ankyrin repeat region 1"
FT Region 157..188
FT /note= "ankyrin repeat region 2"
FT Region 189..222
FT /note= "ankyrin repeat region 3"
XX
XX
EP995798-A1.
XX
XX
PD 26-APR-2000.
XX
XX
PF 24-OCT-1998; 98EP-00120095.
XX
XX
PR 24-OCT-1998; 98EP-00120095.

XX
PA (NOVI-) NOVIIMUNE SA.
XX Masternak K, Reith W, Mach B;
XX
XX WPI; 2000-294958/26.
DR N-PSDB; AA294868.
XX
XX
PT Novel isolated transcription factor, RFXANK, useful for treating MHC
PT class II deficiency and autoimmune disorders, e.g. insulin dependent
PT diabetes and multiple sclerosis, restores the functional transcription of
PT MHC class II genes.
XX
XX
PS Claim 1; Fig 3; 48pp; English.
XX
XX
CC The present sequence is that of human RFXANK, a novel transcription
CC factor that is a subunit of the RFX heterotrimeric transcription complex
CC that binds to the conserved X box motif of all MHC class II gene
CC promoters. The RFXANK gene is mutated in complementation group B MHC II
CC deficiency patients. Mutations identified in patients include aberrant
CC splicing and short deletions in exon 6. The invention provides inhibitors
CC of RFXANK including antibodies, single chain antibodies, dominant
CC negative mutants, antisense molecules and ribozymes. The inhibitors may
CC be used in therapy or prevention of diseases associated with aberrant
CC expression of MHC class II genes and/or as an immunosuppressive agent,
CC e.g. to treat inflammation, autoimmune diseases or rejection of
CC transplanted organs, insulin dependent diabetes, multiple sclerosis,
CC lupus erythematosus and rheumatoid arthritis. The compositions may also
CC be used to treat the autosomal recessive disease MHC class II deficiency.
CC Since RFXANK does not play any other major role in the transcriptional
CC control of genes other than MHC class II genes, its inhibitors are devoid
CC of other undesirable inhibitory effects
XX
XX
SQ Sequence 260 AA;

Query Match 100.0%; Score 1341; DB 3; Length 260;
Best Local Similarity 100.0%; Pred. No. 5.9e-128;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELTQPAEDLIQTQTPASELGDPEDEGAEADGSDTVVLSLFPCTPEPVNPEPDASVSS 60
DB 1 MELTQPAEDLIQTQTPASELGDPEDEGAEADGSDTVVLSLFPCTPEPVNPEPDASVSS 60
QY 61 PQAGSLKHSTTLTNQRGNEVSALPATLDSLSIHQAAQGLDQKHELRKGDNLVNP 120
DB 61 PQAGSLKHSTTLTNQRGNEVSALPATLDSLSIHQAAQGLDQKHELRKGDNLVNP 120
QY 121 DERGFPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERD 180
DB 121 DERGFPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERD 180
QY 181 VDINIYDNGGTPLLYAVRGNHVKCVALLARGADLTTEADSGYTPMDLAVALGYRKVQ 240
DB 181 VDINIYDNGGTPLLYAVRGNHVKCVALLARGADLTTEADSGYTPMDLAVALGYRKVQ 240
QY 241 VIENHILKLFQSNLVPADPE 260
DB 241 VIENHILKLFQSNLVPADPE 260

RESULT 3
ADD27859
ID ADD27859 standard; protein; 260 AA.
XX
XX
AC ADD27859;
XX
XX
DT 15-JAN-2004 (first entry)
XX
XX
DE Human ankyrin family protein (ANFP).
XX
XX
KW Human; ankyrin family protein; ANFP; cell growth;
KW altered cell morphology; NIH3T3 cell; autoimmune disorder;
KW inflammatory disorder; cell proliferative disorder;

KW vesicle trafficking disorder; immunosuppressive; antiinflammatory;
 KW anti-HIV; anti-allergic; antianemic; antiasthmatic; antiarteriosclerotic;
 KW antithyroid; thymometric; neuroprotective; osteopathic; antiarthritic;
 KW antirheumatic; dermatological; haemostatic; antitumor; virucide;
 KW hepatotropic; antipsoriatic; cytostatic; antidiabetic; antilipaemic.
 XX
 OS Homo sapiens.
 XX
 XX US6590077-B1.
 XX
 PD 08-JUL-2003.
 XX
 XX 23-SEP-1999; 99US-00404108.
 XX
 XX 14-OCT-1998; 98US-00172977.
 XX
 XX (INCY-) INCYTE CORP.
 XX
 XX Tang YT, Guegler KJ, Corley NC, Yue H;
 XX WPI; 2003-810387/76.
 XX N-PSDB; ADD27860.
 XX
 XX New human ankyrin family protein, for diagnosing, treating or preventing
 PT autoimmune or inflammatory disorders (e.g. acquired immunodeficiency
 PT syndrome, or multiple sclerosis), hepatitis, leukemia, brain cancer, or
 PT diabetes.
 PT
 PT
 XX
 PS Claim 1; Fig 1A-ID; 35pp; English.
 XX
 CC The present invention relates to the isolation of human ankyrin family
 CC protein (ANFP), and the polynucleotide sequence encoding it. Human ANFP
 CC stimulates increased growth or altered cell morphology in NIH3T3 cells.
 CC The polypeptide is useful for diagnosing, treating or preventing
 CC disorders associated with the expression of ANFP. These disorders include
 CC autoimmune or inflammatory disorders (e.g. AIDS, Addison's disease, adult
 CC respiratory distress syndrome, allergies, anaemia, asthma,
 CC atherosclerosis, autoimmune thyroiditis, Crohn's disease, Graves'
 CC disease, Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis,
 CC pancreatitis, rheumatoid arthritis, scleroderma, Sjogren's syndrome,
 CC systemic lupus erythematosus, thrombocytopenic purpura, ulcerative
 CC colitis), cell proliferative disorders (e.g. actinic keratosis,
 CC arteriosclerosis, bursitis, hepatitis, polycythemia vera, psoriasis,
 CC leukaemia, lymphoma, melanoma, myeloma, sarcoma, bone cancer, breast
 CC cancer, brain cancer), or vesicle trafficking disorders (e.g. cystic
 CC fibrosis, glucose-galactose malabsorption syndrome,
 CC hypercholesterolaemia, diabetes, hyper- or hypoglycaemia, goiter,
 CC ulcers). The present sequence represents human ANFP.
 XX
 SQ Sequence 260 AA;
 Query Match 100.0%; Score 1341; DB 7; Length 260;
 Best Local Similarity 100.0%; Pred. No. 5.9e-128; Indels 0; Gaps 0;
 Matches 260; Conservative 0; Mismatches 0;
 QY 1 MELTPAEDLIQTQTPASELGDPEDPGEAAGSDTVVLSLFPCTPVPNPEPDASVSS 60
 DB 1 MELTPAEDLIQTQTPASELGDPEDPGEAAGSDTVVLSLFPCTPVPNPEPDASVSS 60
 QY 61 PQAGSLKHSTTLTNRQGNVSALPATLDSLSIHQLAAQGLDQKEHLRGDNLVKNP 120
 DB 61 PQAGSLKHSTTLTNRQGNVSALPATLDSLSIHQLAAQGLDQKEHLRGDNLVKNP 120
 QY 121 DERGFPLTWASAFGEIETVRFLEWGDADPHILAKERSALSLSASTGGYTDIVGLLLERD 180
 DB 121 DERGFPLTWASAFGEIETVRFLEWGDADPHILAKERSALSLSASTGGYTDIVGLLLERD 180
 QY 181 VDINIYDNGGTPLLYAVRGNHVKVCEALLIARGADLTTEADSGYTPMDLAVGKRVQ 240
 DB 181 VDINIYDNGGTPLLYAVRGNHVKVCEALLIARGADLTTEADSGYTPMDLAVGKRVQ 240
 QY 241 VIENHILKLFQSNLVPADPE 260
 ||||||||||||||||||

DB 241 VIENHILKLFQSNLVPADPE 260
 RESULT 4
 AAY59590
 ID AAY59590 standard; protein; 269 AA.
 XX
 AC AAY59590;
 XX
 DT 03-APR-2000 (first entry)
 XX
 DE Ankyrin repeat protein TVL-1.
 XX
 XX Ankyrin repeat protein; TVL-1; TNF; tumour necrosis factor; apoptosis;
 KW cell cycle regulation; apoptotic cell death; cell proliferation.
 XX
 OS Mus sp.
 XX
 XX WO9967269-A1.
 XX
 PD 29-DEC-1999.
 XX
 XX 24-JUN-1999; 99WO-US014353.
 XX
 XX 24-JUN-1998; 98US-0090742P.
 XX
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 XX
 XX Tsichlis PN, Makris A;
 XX WPI; 2000-106276/09.
 DR N-PSDB; AAZ49069.
 XX
 PT Novel isolated nucleic acid useful for biological screens to identify
 PT therapeutic agents involved in regulation of cell cycle progression and
 PT apoptotic cell death.
 XX
 PS Claim 12; Fig 2C; 112pp; English.
 XX
 CC This sequence is the ankyrin repeat protein, TVL-1 of the invention. The
 CC protein contains ankyrin repeat domains and is capable of promoting TNF
 CC (tumour necrosis factor) induced apoptosis. The nucleic acid molecules,
 CC proteins and antibodies are useful as targets for screening therapeutic
 CC agents that regulate cell cycle progression and apoptotic cell death,
 CC especially useful for identification, detection and/or regulation of
 CC complex signalling events that regulate cell cycle progression and
 CC apoptotic cell death. The tvl-1 molecules of the invention can also be
 CC used as a research tool and will facilitate the elucidation of the
 CC mechanistic action of the novel genetic and protein interactions involved
 CC in the control of cellular proliferation and apoptosis
 XX
 SQ Sequence 269 AA;
 Query Match 83.0%; Score 1113.5; DB 3; Length 269;
 Best Local Similarity 81.9%; Pred. No. 9.7e-105; Indels 11; Gaps 2;
 Matches 221; Conservative 15; Mismatches 23;
 QY 1 MELTPAEDLIQTQTPASELGDPEDPGEAAGSDTVVLSLFPCTPVPNPEPDASVSS 60
 DB 1 MELTPAEDLIQTQTPASELGDPEDPGEAAGSDTVVLSLFPCTPVPNPEPDASVSS 60
 QY 61 PQAGSLKHSTTLTNRQGNVSALPATLDSLSIHQLAAQGLDQKEHLRK----- 112
 DB 61 LQ-GSFLKHSHTLTNRQGNVSALPATLDSLSIHQLAAQGLDQKEHLRK----- 119
 QY 113 --GDMVKNKPDGERTPLIMASAFGEIETVRFLEWGDADPHILAKERSALSLSASTGGYT 170
 DB 120 LSGNNLKNKPDGERTPLIMASAFGEIETVRFLEWGDADPHILAKERSALSLSASTGGYT 179
 QY 171 DIVGLLLERDVIDINTYDNGGTPLLYAVRGNHVKVCEALLIARGADLTTEADSGYTPMDIA 230
 DB 180 DIVRLLLRDVIDINTYDNGGTPLLYAVRGNHVKVCEALLIARGADLTTEADSGYTPMDIA 239
 ||||||||||||||||||

Query Match 46.3%; Score 621.5; DB 4; Length 313;
 Best Local Similarity 60.6%; Pred. No. 1.8e-54;
 Matches 126; Conservative 26; Mismatches 49; Indels 7; Gaps 2;

QY 51 NPEPDASV-----SSPQAGSSLKHSHTLTNRQGNVSALPATLDSLSIHQAAQGLD 104
 DB 104 SPSPGIQVRHVYTPSTTKHFSPIKQSTTLTNKRGNEVSTTPLLANSLSVHQAAQGLM 163
 QY 105 QLKEHLRKGDNLVNKPDERGFTPLIWASAFGEIETVRFLLEWGADPHILAKERESALS 164
 DB 164 YLATRIEQ-ENVINHDEEGFTPLMWAARHGQIAVVEFLQNGADPQLLGKGRESALS 222
 QY 165 STGGYTDIVGLLLERDVINIDYDNGGTPLLYAVRGNHVKVCEALLARGADLTTEADSGY 224
 DB 223 CSKGYTDIVKMLDCGVNDYDNGGTPLLYAVHGNHVKVCMLESGADPTIETDSGY 282
 QY 225 TPDMLAVALGYRKVQVQVNIENHILKLFQS 252
 DB 283 NSMDLAVALGYRSVQVQVIESHLLKULQN 310

RESULT 7
 AAB94322
 AC AAB94322 standard; protein; 313 AA.
 AC AAB94322;
 DT 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:14803.
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 PN EP1074617-A2.
 PD 07-FEB-2001.
 PF 28-JUL-2000; 2000EP-00116126.
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 PA (HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 14803; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in the CC specification. The primer sets can be used in antisense therapy and in CC gene therapy. The primers are useful for synthesizing polynucleotides, CC particularly full-length cDNAs. The primers are also useful for the CC detection and/or diagnosis of the abnormality of the proteins encoded by CC the full-length cDNAs. The primers allow obtaining of the full-length CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent CC oligonucleotides, all of which are used in the exemplification of the CC present invention
 XX
 SQ Sequence 313 AA;

Query Match 46.3%; Score 621.5; DB 4; Length 313;
 Best Local Similarity 60.6%; Pred. No. 1.8e-54;
 Matches 126; Conservative 26; Mismatches 49; Indels 7; Gaps 2;

QY 51 NPEPDASV-----SSPQAGSSLKHSHTLTNRQGNVSALPATLDSLSIHQAAQGLD 104
 DB 104 SPSPGIQVRHVYTPSTTKHFSPIKQSTTLTNKRGNEVSTTPLLANSLSVHQAAQGLM 163
 QY 105 QLKEHLRKGDNLVNKPDERGFTPLIWASAFGEIETVRFLLEWGADPHILAKERESALS 164
 DB 164 YLATRIEQ-ENVINHDEEGFTPLMWAARHGQIAVVEFLQNGADPQLLGKGRESALS 222
 QY 165 STGGYTDIVGLLLERDVINIDYDNGGTPLLYAVRGNHVKVCEALLARGADLTTEADSGY 224
 DB 223 CSKGYTDIVKMLDCGVNDYDNGGTPLLYAVHGNHVKVCMLESGADPTIETDSGY 282
 QY 225 TPDMLAVALGYRKVQVQVNIENHILKLFQS 252
 DB 283 NSMDLAVALGYRSVQVQVIESHLLKULQN 310

RESULT 8
 ADE60905
 ID ADE60905 standard; protein; 314 AA.
 AC ADE60905;
 DT 29-JAN-2004 (first entry)
 DE Human Protein AK022876, SEQ ID NO 6819.
 KW Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 OS Homo sapiens.
 PN WO2003016475-A2.
 PD 27-FEB-2003.
 PF 14-AUG-2002; 2002WO-US025765.
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-033347P.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.

Woolf C, D'urso D, Befort K, Costigan M;
 WPI; 2003-268312/26.
 GENBANK; AK022876.

New composition comprising two or more isolated polypeptides, useful for preparing a medicament for treating pain in an animal.

Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
CCC or human polynucleotides or a polynucleotide which represents a fragment,
CCC derivative or allelic variation of the nucleic acid sequence. Also
CCC claimed are a vector comprising the novel polynucleotide, a host cell
CCC comprising the vector, a method for identifying a nucleotide sequence
CCC which is differentially regulated in an animal subjected to pain and a
CCC kit to perform the method, an array, a method for identifying an agent
CCC that increases or decreases the expression of the polynucleotide sequence
CCC that is differentially expressed in neuronal tissue of a first animal
CCC subjected to pain, a method for identifying a compound which regulates
CCC the expression of a polynucleotide sequence which is differentially
CCC expressed in an animal subjected to pain, a method for identifying a
CCC compound that regulates the activity of one or more of the
CCC polynucleotides, a method for producing a pharmaceutical composition, a
CCC method for identifying a compound or small molecule that regulates the
CCC activity in an animal of one or more of the polypeptides given in the
CCC specification, a method for identifying a compound useful in treating
CCC pain and a pharmaceutical composition comprising the one or more
CCC polypeptides or their antibodies. The polynucleotide or the compound that
CCC modulates its activity is useful for preparing a medicament for treating
CCC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CCC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CCC therapy). The sequence presented is a human protein (shown in Table 2 of
CCC the specification) which is differentially expressed during pain. Note:
CCC The sequence data for this patent did not form part of the printed
CCC specification, but was obtained in electronic form directly from WIPO at
CCC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 314 AA:
SO

[illegible]

14-AUG-2001; 2001US-0312147P.
01-NOV-2001; 2001US-0346382P.
26-NOV-2001; 2001US-0333347P.
(GEOH) GEN HOSPITAL CORP.
(FARB) BAYER AG.
Wolff C, D'urso D, Befort K, Costigan M;
WPI; 2003-2669312/26.
GENBANK; AK022876.
New composition comprising two or more isolated polypeptides, useful for
preparing a medicament for treating pain in an animal.
Claim 1; Page; 1017pp; English.
The invention discloses a composition comprising two or more isolated rat
or human polynucleotides or a polynucleotide which represents a fragment,
derivative or allelic variation of the nucleic acid sequence. Also
claimed are a vector comprising the novel polynucleotide, a host cell
comprising the vector, a method for identifying a nucleotide sequence
which is differentially regulated in an animal subjected to pain and a
kit to perform the method, an array, a method for identifying an agent
that increases or decreases the expression of the polynucleotide sequence
that is differentially expressed in neuronal tissue of a first animal
subjected to pain, a method for identifying a compound which regulates
the expression of a polynucleotide sequence which is differentially
expressed in an animal subjected to pain, a method for identifying a
compound that regulates the activity of one or more of the
polynucleotides, a method for producing a pharmaceutical composition, a
method for identifying a compound or small molecule that regulates the
activity in an animal of one or more of the polypeptides given in the
specification, a method for identifying a compound useful in treating
pain and a pharmaceutical composition comprising the one or more
polypeptides or their antibodies. The polynucleotide or the compound that
modulates its activity is useful for preparing a medicament for treating
pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
therapy). The sequence presented is a human protein (shown in Table 2 of
the specification) which is differentially expressed during pain. Note:
The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic form directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.
Sequence 314 AA:
XX
SO

Query Match	46.3%	Score	621.5	DB	7	Length	314
Best Local Similarity	60.6%	Pred. No.	1.8e-54				
Matches	126	Conservative	26	Mismatches	49	Indels	7
Gaps							
QY	51	NPEFDASV-----SPQAGSSLKHSHTLTNRORGNEVSALPATLDSLSIHQLAAQ	GELD	104			
Db	104	SPSPGIQVRHVYTPSTTKHSPKIQSTLTLTNKHGNEVSTPTLLANSLSVHQLAAQ	GEML	163			
QY	105	QLKEHLRKGDNLVKNKPDERGFTPLIIVASAFGEIETVRFLEWGADPHILAKERSALS	LSLA	164			
Db	164	YLATRIEQ-ENVINHDTDEGFTPLMWAHGHQIAVVEFLQNGADPQLCKGRESALS	LSLA	222			
QY	165	STCGYTDVIGLLERDVIDINYPWNGTPTLLYAVRGNHVKCVKALLARGADLTTEAD	SGY	224			
Db	223	CSKGYTDIVKMLDCCGVNVEYDWNNGTPTLLYAVHGNHVKCVKMLLES	GADPTIETDSG	282			
QY	225	TPMDLVALGYRKVQVQVIENHILKLFQS	252				
Db	283	NSMDLVALGYRSVQVQVIESHLKLQ	LN	310			
RESULT	10						
ADD47704							
ID	ADD47704	standard	protein	312	AA		
XX							

PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 06-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.

PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-451931/48.
 XX N-PSDB; AAS33355.
 DR
 XX
 PT New nucleic acids and polypeptides, useful for diagnosing, preventing or
 PT treating medical conditions.
 XX
 XX Claim 11; SEQ ID NO 638; 753pp; English.

CC The invention relates to novel isolated nucleic acid molecules (I)
 CC encoding human secreted proteins (II). (I) and (II) are used to prevent,
 CC treat or ameliorate a medical condition in e.g. humans, mice, rabbits, in
 CC goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in
 CC the prevention, treatment and diagnosis of diseases associated with
 CC inappropriate expression of secreted proteins. (I) and complementary
 CC sequences may also be used as DNA probes in diagnostic assays (e.g.
 CC polymerase chain reactions (PCR)) to detect and quantitate the presence
 CC of similar nucleic acid sequences in samples, and so which patients may
 CC be in need of restorative therapy. (II) may also be used as antigens in
 CC the production of antibodies and in assays to identify modulators
 CC (agonists and antagonists) of the expression and activity of the secreted
 CC proteins. The anti-(II) antibodies and antagonists may also be used to
 CC down regulate expression and activity of (II). The anti-(II) antibodies
 CC may also be used as diagnostic agents for detecting the presence of (II)
 CC in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The
 CC disorders include for example: immune/autoimmune diseases (e.g. HIV
 CC (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
 CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
 CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
 CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
 CC Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/
 CC cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and
 CC thrombosis), infections caused by bacteria, viruses and fungi and ocular
 CC disorders (e.g. corneal infections). (I) and (II), agonists, antagonists
 CC and antibodies can also be used to promote wound healing, maintain organs
 CC before transplantation, and support cell culture of primary tissues.

Query Match 45.3%; Score 607.5; DB 4; Length 263;
 Best Local Similarity 59.4%; Pred. No. 3.7e-53;
 Matches 123; Conservative 26; Mismatches 51; Indels 7; Gaps 2;

QY 1 MELTQPAEDLIQTQTPASBELGDPGEAAAGSDTVVLSPFCTPEPVNPEPDASVSS 60
 Db 1 MELTQPAEDLIQTQTPASBELGDPGEAAAGSDTVVLSPFCTPEPVNPEPDASVSS 60
 QY 61 PQAGSSIKHSTTLTNQRGNVSAL 85
 Db 61 PQ-GSSIKHSTTLTNQRGNVSAL 84

RESULT 14

AAU20665

ID AAU20665 standard; protein; 119 AA.

XX

AC AAU20665;

XX

DT 04-DEC-2001 (first entry)

XX

DE Human secreted protein, Seq ID No 657.

XX

KW Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;
 KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;
 KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;
 KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;
 KW multiple sclerosis; cancer; hyperproliferative disorder; infection;
 KW Gaucher's disease; neurological disease; cerebrovascular disorder;
 KW thrombosis; wound healing.

XX

OS Homo sapiens.

XX

PN WO20015326-A2.

XX

PD 02-AUG-2001.

XX

XX 17-JAN-2001; 2001WO-US001347.

XX

PR 31-JAN-2000; 2000US-0179065P.

PR

PR 04-FEB-2000; 2000US-0180628P.

PR

PR 24-FEB-2000; 2000US-0184664P.

PR

PR 02-MAR-2000; 2000US-0186350P.

PR

PR 16-MAR-2000; 2000US-0189874P.

PR

PR 17-MAR-2000; 2000US-0190076P.

PR

PR 18-APR-2000; 2000US-0198123P.

PR

PR 19-MAY-2000; 2000US-0205515P.

PR

PR 07-JUN-2000; 2000US-0209467P.

PR

PR 28-JUN-2000; 2000US-0214886P.

PR

PR 30-JUN-2000; 2000US-0215135P.

PR

PR 07-JUL-2000; 2000US-0216647P.

PR

PR 07-JUL-2000; 2000US-0216860P.

PR

PR 11-JUL-2000; 2000US-0217487P.

PR

PR 14-JUL-2000; 2000US-0217496P.

PR

PR 26-JUL-2000; 2000US-0220963P.

PR

PR 26-JUL-2000; 2000US-0220964P.

PR

PR 14-AUG-2000; 2000US-0224518P.

PR

PR 14-AUG-2000; 2000US-0224519P.

PR

PR 14-AUG-2000; 2000US-0225213P.

PR

PR 14-AUG-2000; 2000US-0225214P.

PR

PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 12-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241421P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 17-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.

Db 65 DPTIETDSGYNMDLAVALGYRSVQQVIESHLLKLQN 102

Search completed: April 29, 2004, 06:25:49
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 06:26:53 ; Search time 47 Seconds
(without alignments)

1533.389 Million cell updates/sec

Title: US-09-840-243C-11

Perfect score: 1341

Sequence: 1 MELTQPAEDLIQTQTPASE.....VIENHILKLFQSNLVPADPE 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1341	100.0	260	9	US-09-840-243B-11
2	1341	100.0	260	9	US-09-840-243B-12
3	1113.5	83.0	269	9	US-09-840-243B-13
4	625	46.6	229	15	US-10-264-049-3013
5	561.5	41.9	220	9	US-09-840-243B-18
6	542.5	40.5	218	9	US-09-840-243B-19
7	495	36.9	152	12	US-10-221-625-12
8	257	19.2	49	9	US-09-864-761-47859
9	216.5	16.1	747	15	US-10-094-749-1924
10	212	15.8	705	14	US-10-106-698-6378
11	212	15.8	1184	14	US-10-117-229-3
12	212	15.8	1184	14	US-10-117-229-9
13	212	15.8	1762	14	US-10-117-229-7
14	212	15.8	1771	14	US-10-117-229-4
15	212	15.8	1771	14	US-10-117-229-11

16	207.5	15.5	1715	14	US-10-021-571-2	Sequence 2, Appli
17	207.5	15.5	1715	14	US-10-021-571-4	Sequence 4, Appli
18	207.5	15.5	1715	14	US-10-117-229-2	Sequence 2, Appli
19	207.5	15.5	1715	14	US-10-117-229-6	Sequence 6, Appli
20	202.5	15.1	720	16	US-10-433-794-20	Sequence 20, Appli
21	202.5	15.1	765	12	US-10-182-243-56	Sequence 56, Appli
22	200.5	15.0	301	12	US-10-112-944-738	Sequence 738, App
23	200.5	15.0	301	12	US-10-112-944-739	Sequence 739, App
24	200.5	15.0	301	12	US-10-112-944-740	Sequence 740, App
25	199.5	14.9	657	15	US-10-104-047-2529	Sequence 2529, Ap
26	199	14.8	1498	14	US-10-021-571-8	Sequence 8, Appli
27	198.5	14.8	765	14	US-10-128-174-3	Sequence 3, Appli
28	198.5	14.8	765	14	US-10-128-174-34	Sequence 34, Appli
29	198.5	14.8	765	14	US-10-128-174-35	Sequence 35, Appli
30	198.5	14.8	765	14	US-10-128-174-36	Sequence 36, Appli
31	198.5	14.8	765	14	US-10-128-174-37	Sequence 37, Appli
32	198.5	14.8	765	14	US-10-128-174-38	Sequence 38, Appli
33	198.5	14.8	765	14	US-10-128-174-39	Sequence 39, Appli
34	198.5	14.8	765	14	US-10-128-174-40	Sequence 40, Appli
35	198.5	14.8	765	14	US-10-128-174-41	Sequence 41, Appli
36	198.5	14.8	765	14	US-10-128-174-42	Sequence 42, Appli
37	198.5	14.8	765	14	US-10-128-174-43	Sequence 43, Appli
38	198.5	14.8	765	14	US-10-128-174-44	Sequence 44, Appli
39	198.5	14.8	1330	15	US-10-108-260A-3237	Sequence 3237, Ap
40	196.5	14.7	551	9	US-09-835-788A-17	Sequence 17, Appl
41	196.5	14.7	551	9	US-10-175-042-17	Sequence 17, Appl
42	192	14.3	740	9	US-09-835-788A-12	Sequence 12, Appl
43	192	14.3	740	14	US-10-175-042-12	Sequence 12, Appl
44	191	14.2	426	9	US-09-908-711-70	Sequence 70, Appl
45	190	14.2	309	12	US-10-425-114-54227	Sequence 54227, A

ALIGNMENTS

RESULT 1

US-09-840-243B-11
; Sequence 11, Application US/09840243B
; Patent No. US20020156258A1
; GENERAL INFORMATION:
; APPLICANT: MASTERNAK, Krzysztof
; APPLICANT: REITH, Walter
; APPLICANT: MACH, Bernard
; TITLE OF INVENTION: New Transcription Factor of MHC Class II Genes, Substances
; TITLE OF INVENTION: Capable of Inhibiting This New Transcription Factor and
; TITLE OF INVENTION: Medical Uses of These Substances
; FILE REFERENCE: 010830-117
; CURRENT APPLICATION NUMBER: US/09/840,243B
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: EP 98120085.0
; PRIOR FILING DATE: 1998-10-24
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-840-243B-11

Query Match 100.0%; Score 1341; DB 9; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.1e-115;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MELTQPAEDLIQTQTPASELGDPEGEAAGSDGTWVLSLFPCTPEPVNPEPDASVSS	60
DB	1	MELTQPAEDLIQTQTPASELGDPEGEAAGSDGTWVLSLFPCTPEPVNPEPDASVSS	60
QY	61	FOAGSLKHSTLTNRQRNEVSALPATLDSLSLHQLAAQGELOLKEHLRKGDNLVNKP	120
DB	61	FOAGSLKHSTLTNRQRNEVSALPATLDSLSLHQLAAQGELOLKEHLRKGDNLVNKP	120
QY	121	DERGTPILWASAFGEIETVRFLEBWGADPHILAKERESALSTGGYTTIVGLLLERD	180

Db 121 DERGFTPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERD 180
Qy 181 VDINIYDNGGTPLLIYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVAGYRKVQ 240
Db 181 VDINIYDNGGTPLLIYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVAGYRKVQ 240
Qy 241 VIENHILKLFOSNLVPADPE 260
Db 241 VIENHILKLFOSNLVPADPE 260

RESULT 2

US-09-840-243B-12
; Sequence 12, Application US/09840243B
; Patent No. US20020156258A1
; GENERAL INFORMATION:
; APPLICANT: MASTERNAK, Krzysztof
; APPLICANT: REITH, Walter
; APPLICANT: MACH, Bernard
; TITLE OF INVENTION: New Transcription Factor of MHC Class II Genes, Substances
; TITLE OF INVENTION: Capable of Inhibiting This New Transcription Factor and
; TITLE OF INVENTION: Medical Uses of These Substances
; FILE REFERENCE: 010830-117
; CURRENT APPLICATION NUMBER: US/09/840,243B
; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: EP 98120085.0
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-840-243B-12

Query Match 100.0%; Score 1341; DB 9; Length 260;
Best Local Similarity 100.0%; Pred. No. 1.1e-115;
Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MELTPQAEADLIQTOOTPASELGDPDPGEAAGDSVTWLSLFPCTPEPVNPEPDASVSS 60
Db 1 MELTPQAEADLIQTOOTPASELGDPDPGEAAGDSVTWLSLFPCTPEPVNPEPDASVSS 60
Qy 61 PQAGSSLKHSHTLTNRQGNVSALPATLDSLSIHQAAQAGELDQKHELRKGDNLVNP 120
Db 61 PQAGSSLKHSHTLTNRQGNVSALPATLDSLSIHQAAQAGELDQKHELRKGDNLVNP 120
Qy 121 DERGFTPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERD 180
Db 121 DERGFTPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLERD 180
Qy 181 VDINIYDNGGTPLLIYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVAGYRKVQ 240
Db 181 VDINIYDNGGTPLLIYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVAGYRKVQ 240
Qy 241 VIENHILKLFOSNLVPADPE 260
Db 241 VIENHILKLFOSNLVPADPE 260

RESULT 3

US-09-840-243B-13
; Sequence 13, Application US/09840243B
; Patent No. US20020156258A1
; GENERAL INFORMATION:
; APPLICANT: MASTERNAK, Krzysztof
; APPLICANT: REITH, Walter
; APPLICANT: MACH, Bernard
; TITLE OF INVENTION: New Transcription Factor of MHC Class II Genes, Substances
; TITLE OF INVENTION: Capable of Inhibiting This New Transcription Factor and
; TITLE OF INVENTION: Medical Uses of These Substances
; FILE REFERENCE: 010830-117
; CURRENT APPLICATION NUMBER: US/09/840,243B

; CURRENT FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: EP 98120085.0
; PRIOR FILING DATE: 1998-10-24
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
US-09-840-243B-13

Query Match 83.0%; Score 1113.5; DB 9; Length 269;
Best Local Similarity 81.9%; Pred. No. 1.3e-94;
Matches 221; Conservative 15; Mismatches 23; Indels 11; Gaps 2;
Qy 1 MELTPQAEADLIQTOOTPASELGDPDPGEAAGDSVTWLSLFPCTPEPVNPEPDASVSS 60
Db 1 MEPTQVAENLVNQPPVDPEDPDRDESPENSMTVLSLFPCTPDVNPEDASASS 60
Qy 61 PQAGSSLKHSHTLTNRQGNVSALPATLDSLSIHQAAQAGELDQKHELRK----- 112
Db 61 LQ-GSFLKHSHTLTNRQGNVSALPATLDSLSIHQAAQAGELSOLKHLRGKACACTC 119
Qy 113 --GDNLVNPDERGFTPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYT 170
Db 120 LSGNNLKNPDERGFTPLIWSAFGEIETVRFLEWGADPHILAKERESALSASTGGYT 179
Qy 171 DIVGLLLERDVIDINIYDNGGTPLLIYAVRGNHVKVEALLARGADLTTEADSGYTPMDLA 230
Db 180 DIVRLLDRDVIDINIYDNGGTPLLIYAVRGNHVKVEALLARGADLTTEADSGYTPMDLA 239
Qy 231 VALGYRKVQVVIENHILKLFQSNLVPADPE 260
Db 240 VALGYRKVQVVMESHILRLFQSTLGFVDPE 269

RESULT 4

US-10-264-049-3013
; Sequence 3013, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 3013
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-264-049-3013

Query Match 46.6%; Score 625; DB 15; Length 229;
Best Local Similarity 58.6%; Pred. No. 1.7e-49;
Matches 130; Conservative 26; Mismatches 58; Indels 8; Gaps 3;
Qy 37 TVVLSLFPCTPEPVNPEPDASV-----SSPQAGSSLKHSHTLTNRQGNVSALPATLD 90
Db 7 TTFVHLAECNIH-TSPSPCIQVRHVYTPSTTKHFSPIKOSTTLTNKRGNEVSTTPLLAN 65
Qy 91 SLSTHQLAAQAGELDQKHELRKGDNLVNPDERGFTPLIWSAFGEIETVRFLEWGADP 150
Db 66 SLSTHQLAAQAGELVLAIRIEQ-ENVINHTDEGFTPLMWAHAHQIAVVEFLQNGADP 124
Qy 151 HILAKERESALSASTGGYTDIVGLLLERDVIDINIYDNGGTPLLIYAVRGNHVKVEALL 210
Db 125 QLLGKGRESALSACSKGYTDIVKMLDCGVNDVNEWNGGTPLLIYAVHGNHVKVRL 184

Qy	234	GYRKVQVVIENHILKLFQS	252
		: :	
Db	131	GYRSVQOVIESHLLKLLON	149

RESULT 8

```

US-09-864-761-47859
; Sequence 47859, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47859
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002126.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EST HUMAN HIT: BE905633.1, EVALUE 3.00e-22
; OTHER INFORMATION: SWISSPROT HIT: Q14593, EVALUE 2.00e-23
US-09-864-761-47859

```

Query Match 19.2%; Score 257; DB 9; Length 49;
Best Local Similarity 100.0%; Pred. No. 2e-16;

	Matches	49; Conservative	0; Mismatches	0; Indels	0; Gaps
QY	1	MELTQPAEDLIQTQTFASLGDPDPGEFAADGSDTVVLSLPCTPEP	49		
Db	1	MELTQPAEDLIQTQTFASLGDPDPGEFAADGSDTVVLSLPCTPEP	49		

RESULT 9

```

US-10-094-749-1924
; Sequence 1924, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NACHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOUYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1924
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1924

```

Query Match	16.1%;	Score 216.5;	DB 15;	Length 747;
Best Local Similarity	27.6%;	Pred. No. 5.8e-11;		
Matches 72: Conservative	49;	Mismatches 99;	Indels 41;	Gaps 9;

QY	2	ELTQP	AEDLIQTOTPASEL---	DDEDGEEAAGSDTVVL-----	SLEFCTP--	47
		:	:	:	:	:
		:	:	:	:	:
Db	165	QLTFPSHDLNRAPRVFVQALPMVVPPQEPDPKPANVATTLPIRNKAVSGRASAMSNTPTH	224			
		:	:	:	:	:
QY	48	-----EPWNPEDASVSPOQSSGLKSTHTLTUNRQRGNVEVALPATLDSLSHQLAQA	100			
		:	:	:	:	:
Db	225	STAASISQOTPTP-SPIISPSAMLFIYPADIDAQTESNHDTAL-----	T-LACA	273		
		:	:	:	:	:
QY	101	GELDQLKEHLRGDNLVNPKPDGERFTPLIWAFAFGIETVRFLLGWGADPHILA-KERES	159			
		:	:	:	:	:
Db	274	GCHEEIVQTLILBERGASIEHRDKKGFTPLIAAVAGHVGVVEILLDNGADIQAOSERTKOT	333			
		:	:	:	:	:
QY	160	ALSLASTGGVTIDVGLLBERDVD--INLYDNWGTTPLLVAVRGNHKVCVEALLARGADL	216			
		:	:	:	:	:
Db	334	PUSLCASGRQEVVELLARGANKERNVSDY---TPLSLAASGGVYNIKIULLNAGAEL	390			
		:	:	:	:	:
QY	217	TTEADS--GYTPMDLVALGY	235			
		:	:	:	:	:
Db	391	NSRTGSKLGISPLMLAAMNGH	411			

RESULT 10
US-10-106-698-6378

```

; Sequence 6378, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6378
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (244)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (337)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6378

Query Match 15.8%; Score 212; DB 14; Length 705;
Best Local Similarity 34.7%; Pred. No. 1.4e-10;
Matches 51; Conservative 33; Mismatches 61; Indels 2; Gaps 2;

QY 100 QGELDOLKEHLKGDNLVKNKPDGFTPLIWSAFGEIETVRFLEWGDADPHILAKERES 159
Db 47 EGHVHIVEELLKCGVNLHR-DMGWTALMWACYKGRDVTVELLSHGANSPTGLYSYV 105

QY 160 ALSLASTGGYTDIVGLLLERDVIDINIDWNGGTPLLIYAVRGHVKVCEALLARGADLTTE 219
Db 106 PIWAAGRGHADIVHLLQNGAKVNCSDKYGTTPLVWAARKHLECVKHLANGADVDOE 165

QY 220 ADSGYTPMDLVALGY-RKVQOVNIENH 245
Db 166 GANSMTALIVAVKGGYTQSVKEILKRN 192

RESULT 11
US-10-117-229-3
; Sequence 3, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-3

Query Match 15.8%; Score 212; DB 14; Length 1184;
Best Local Similarity 34.7%; Pred. No. 3e-10;
Matches 51; Conservative 33; Mismatches 61; Indels 2; Gaps 2;

QY 100 QGELDOLKEHLKGDNLVKNKPDGFTPLIWSAFGEIETVRFLEWGDADPHILAKERES 159
Db 81 EGHVHIVEELLKCGVNLHR-DMGWTALMWACYKGRDVTVELLSHGANSPTGLYSYV 139

QY 160 ALSLASTGGYTDIVGLLLERDVIDINIDWNGGTPLLIYAVRGHVKVCEALLARGADLTTE 219

```

```

; Sequence 6378, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6378
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (244)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (337)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-6378

Query Match 15.8%; Score 212; DB 14; Length 1184;
Best Local Similarity 34.7%; Pred. No. 3e-10;
Matches 51; Conservative 33; Mismatches 61; Indels 2; Gaps 2;

QY 100 QGELDOLKEHLKGDNLVKNKPDGFTPLIWSAFGEIETVRFLEWGDADPHILAKERES 159
Db 81 EGHVHIVEELLKCGVNLHR-DMGWTALMWACYKGRDVTVELLSHGANSPTGLYSYV 139

QY 160 ALSLASTGGYTDIVGLLLERDVIDINIDWNGGTPLLIYAVRGHVKVCEALLARGADLTTE 219
Db 140 PIWAAGRGHADIVHLLQNGAKVNCSDKYGTTPLVWAARKHLECVKHLANGADVDOE 199

QY 220 ADSGYTPMDLVALGY-RKVQOVNIENH 245
Db 200 GANSMTALIVAVKGGYTQSVKEILKRN 226

RESULT 12
US-10-117-229-9
; Sequence 9, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-9

Query Match 15.8%; Score 212; DB 14; Length 1184;
Best Local Similarity 34.7%; Pred. No. 3e-10;
Matches 51; Conservative 33; Mismatches 61; Indels 2; Gaps 2;

QY 100 QGELDOLKEHLKGDNLVKNKPDGFTPLIWSAFGEIETVRFLEWGDADPHILAKERES 159
Db 81 EGHVHIVEELLKCGVNLHR-DMGWTALMWACYKGRDVTVELLSHGANSPTGLYSYV 139

QY 160 ALSLASTGGYTDIVGLLLERDVIDINIDWNGGTPLLIYAVRGHVKVCEALLARGADLTTE 219
Db 140 PIWAAGRGHADIVHLLQNGAKVNCSDKYGTTPLVWAARKHLECVKHLANGADVDOE 199

QY 220 ADSGYTPMDLVALGY-RKVQOVNIENH 245
Db 200 GANSMTALIVAVKGGYTQSVKEILKRN 226

RESULT 13
US-10-117-229-7
; Sequence 7, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1762
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-117-229-7

Query Match 15.8%; Score 212; DB 14; Length 1762;
Best Local Similarity 34.0%; Pred. No. 5.3e-10;
Matches 50; Conservative 34; Mismatches 61; Indels 2; Gaps 2;

QY 100 QGELDOLKEHLKGDNLVKNKPDGFTPLIWSAFGEIETVRFLEWGDADPHILAKERES 159
Db 81 EGHVHIVEELLKCGVNLHR-DMGWTALMWACYKGRDVTVELLSHGANSPTGLYSYV 139

QY 160 ALSLASTGGYTDIVGLLLERDVIDINIDWNGGTPLLIYAVRGHVKVCEALLARGADLTTE 219
Db 140 PIWAAGRGHADIVHLLQNGAKVNCSDKYGTTPLVWAARKHLECVKHLANGADVDOE 199

```


Search completed: April 29, 2004, 06:34:56
Job time : 48 secs

QY 220 ADSGYTPMDLVALGY-RKQVQVIENH 245
Db 200 GANSM TALIVAVKGGYTQSVKEILKRN 226

RESULT 14

US-10-117-229-4
; Sequence 4, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1771
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-4

Query Match 15.8%; Score 212; DB 14; Length 1771;
Best Local Similarity 34.7%; Pred. No. 5.3e-10;
Matches 51; Conservative 33; Mismatches 61; Indels 2; Gaps 2;
QY 100 QGELDQLKEHLKRGDNLVKNKPDGRTPLIWSAFGEIETVRFLEWGDADPHILAKERES 159
Db 81 EGHVHIVEELLKCGVNLEHR-DMGGWTALMWACYKGRDVTVELLSHGANPSVTGLYSYV 139
QY 160 ALSLASTGGYTDIVGLLLERDVIDINIDWNGGTPLLYAVRGNHVKVEALLARGADLTTE 219
Db 140 PIIWAAGRGHADIVHLLQNGAKVNCSDKYGTPTLVWAARKGHLECVKHLAMGADVDOE 199
QY 220 ADSGYTPMDLVALGY-RKQVQVIENH 245
Db 200 GANSM TALIVAVKGGYTQSVKEILKRN 226

RESULT 15

US-10-117-229-11
; Sequence 11, Application US/10117229
; Publication No. US20030190625A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies, Inc.
; TITLE OF INVENTION: Human Kidins220Pc
; FILE REFERENCE: 9U 104 R1
; CURRENT APPLICATION NUMBER: US/10/117,229
; CURRENT FILING DATE: 2002-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1771
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-229-11

Query Match 15.8%; Score 212; DB 14; Length 1771;
Best Local Similarity 34.7%; Pred. No. 5.3e-10;
Matches 51; Conservative 33; Mismatches 61; Indels 2; Gaps 2;
QY 100 QGELDQLKEHLKRGDNLVKNKPDGRTPLIWSAFGEIETVRFLEWGDADPHILAKERES 159
Db 81 EGHVHIVEELLKCGVNLEHR-DMGGWTALMWACYKGRDVTVELLSHGANPSVTGLYSYV 139
QY 160 ALSLASTGGYTDIVGLLLERDVIDINIDWNGGTPLLYAVRGNHVKVEALLARGADLTTE 219
Db 140 PIIWAAGRGHADIVHLLQNGAKVNCSDKYGTPTLVWAARKGHLECVKHLAMGADVDOE 199
QY 220 ADSGYTPMDLVALGY-RKQVQVIENH 245
Db 200 GANSM TALIVAVKGGYTQSVKEILKRN 226

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 06:16:22 ; Search time 17 Seconds
(without alignments)
796.367 Million cell updates/sec

Title: US-09-840-243c-11

Perfect score: 1341

Sequence: 1 MELTQPAEDLITQQTQTPASE.....VIENHILKLFQSLNLPADPE 260

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1341	100.0	260	1	RFXK_HUMAN	O14593 homo sapien
2	1113.5	83.0	269	1	RFXK_MOUSE	Q92205 mus musculus
3	621.5	46.3	313	1	ANRA_HUMAN	Q919e1 homo sapien
4	612.5	45.7	312	1	ANRA_MOUSE	Q919e2 mus musculus
5	206.5	15.4	1267	1	HMT1_HUMAN	Q919b1 homo sapien
6	202.5	15.1	3924	1	ANK2_HUMAN	Q01484 homo sapien
7	193	14.4	4377	1	ANK3_HUMAN	Q12955 homo sapien
8	189	14.1	1210	1	BAT8_HUMAN	Q96kq7 homo sapien
9	188	14.0	1263	1	BAT8_MOUSE	Q92148 mus musculus
10	187	13.9	475	1	GASZ_MOUSE	Q8vd46 mus musculus
11	187	13.9	1327	1	TNK1_HUMAN	Q95271 homo sapien
12	184	13.7	1980	1	ANK1_HUMAN	P16157 homo sapien
13	183	13.6	1862	1	ANK1_MOUSE	Q02357 mus musculus
14	182.5	13.6	231	1	PSDA_MOUSE	Q922x2 mus musculus
15	181.5	13.5	768	1	YB23_HUMAN	Q9ulj7 homo sapien
16	180	13.4	475	1	GASZ_PANTR	Q8wmx6 pan troglod
17	179	13.3	475	1	GASZ_HUMAN	Q8whh4 homo sapien
18	176.5	13.2	226	1	PSDA_HUMAN	Q75832 homo sapien
19	176.5	13.2	231	1	PSDA_RAT	Q922x3 rattus norv
20	175	13.0	328	1	ANK2_MOUSE	Q9ww06 mus musculus
21	175	13.0	435	1	AKR_ARATH	Q05753 arabidopsis
22	174.5	13.0	360	1	Y045_METWA	Q8g0u0 methanosarc
23	173	12.9	333	1	ANK2_HUMAN	Q9qzv1 homo sapien
24	173	12.9	415	1	ANKX_MOUSE	Q991w0 mus musculus
25	173	12.9	1150	1	ANK1_MOUSE	P59672 mus musculus
26	172.5	12.9	1059	1	Y379_HUMAN	Q15084 homo sapien
27	171.5	12.8	420	1	ANKX_HUMAN	P9nxr5 homo sapien
28	169.5	12.6	451	1	ILK_CAVPO	P57044 cavia porce
29	169.5	12.6	741	1	RN5A_HUMAN	Q05823 homo sapien
30	169	12.6	256	1	Y161_PYRAE	Q82wc4 pyrobaculum
31	168	12.5	776	1	ANK5_HUMAN	Q9nu02 homo sapien
32	166.5	12.4	1442	1	DAK1_MOUSE	Q80ye7 mus musculus
33	165.5	12.3	452	1	ILK1_HUMAN	Q13418 homo sapien

RESULT 1

ID	RFXK_HUMAN	STANDARD;	PRT;	260 AA.
AC	O14593; O95839;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	DNA-binding protein RFXANK (Regulatory factor X subunit B) (RFX-B)			
DE	(Ankyrin repeat family A protein 1).			
GN	RFXANK OR RFXB OR ANKRA1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND MASS SPECTROMETRY.			
RC	TISSUE=B-cell;			
EX	MEDLINE=99021383; PubMed=9806546;			
RA	Masternak K., Barras E., Zufferey M., Conrad B., Cortals G.,			
RA	Aebersold R., Sanchez J.-C., Hochstrasser D.F., Mach B., Reith W.;			
RT	"A gene encoding a novel RFX-associated transactivator is mutated in			
RT	the majority of MHC class II deficiency patients.";			
RL	Nat. Genet. 20:273-277(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 79-95; 180-198; 100-210 AND			
RC	TISSUE=Lymphoblast;			
EX	MEDLINE=99170284; PubMed=10072068;			
RA	Nagarajan U.M., Louis-Pence P., DeSandro A., Nilsen R., Bushey A.,			
RA	Boss J.M.;			
RT	RFX-B is the gene responsible for the most common cause of the bare			
RT	lymphocyte syndrome, an MHC class II immunodeficiency.";			
RL	Immunity 10:153-162(1999).			
RN	[3]			
RP	ERRATUM.			
RA	Nagarajan U.M., Louis-Pence P., DeSandro A., Nilsen R., Bushey A.,			
RA	Boss J.M.;			
RL	Immunity 10:399-399(1999).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Zhou J., Guan Z., Gu J., Ye M., Fu G., Zhang Q., Xu S., He K.,			
RA	Chen S., Mao M., Chen Z.;			
RT	"Hematopoietic-derived ankyrin-like gene.";			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE OF 65-260 FROM N.A.			
RA	Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,			
RA	Christensen M., Gordon L., Kyle A., Ramirez M., Stilwagen S.,			
RA	Garnes J., Danganan L., Bruce R., Quan G., Montgomery M., Ow D.,			
RA	Kobayashi A., Olsen A.O., Carrano A.V.;			
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	VARIANT BLS II GROUP B PRO-195.			
RA	MEDLINE=20192025; PubMed=10725724;			
RA	Nagarajan U.M., Fejtenburg A., Gobin S.J., Boss J.M.,			
RA	van den Eijzen P.J.;			

ALIGNMENTS

34	165.5	12.3	452	1	ILK2_HUMAN	P57043 homo sapien
35	164	12.2	642	1	YA2A_SCHPO	Q09701 schizosacch
36	163.5	12.2	518	1	ASB3_HUMAN	Q95751 homo sapien
37	163	12.2	117	1	MTFN_MOUSE	P80144 mus musculus
38	163	12.2	118	1	MTFN_CHICK	Q91955 gallus gall
39	163	12.2	1401	1	LATA_LATWA	P23631 latrodectus
40	162.5	12.1	1134	1	ANS1_HUMAN	Q92625 homo sapien
41	162	12.1	656	1	FEM1_CAEEL	P17221 caenorhabdi
42	161	12.0	117	1	MTFN_HUMAN	P58546 homo sapien
43	161	12.0	1432	1	DAK1_HUMAN	P53355 homo sapien
44	160.5	12.0	452	1	ILK_MOUSE	O55222 mus musculus
45	160.5	12.0	525	1	ASB3_MOUSE	Q9wv72 mus musculus

"Novel mutations within the RFX-B gene and partial rescue of MHC and related genes through exogenous class II transactivator in RFX-B-deficient cells."

CC J. Immunol. 164:3666-3674 (2000).

CC -!- FUNCTION: Activates transcription from class II MHC promoters. Activation requires the activity of the MHC class II transactivator (MHC2TA). May regulate other genes in the cell. RFX binds the X1 box of MHC-II promoters. Isoform RFX-B-DELTA5 is not involved in the positive regulation of MHC class II genes.

CC -!- SUBUNIT: Rfx consists of at least three different subunits; RFXAP, RFX5 and RFX-B/REFXANK; with each subunit representing a separate complementation group. RFX forms cooperative DNA binding complexes with X2BP and CBF/NF-Y. RFX associates with MHC2TA to form an active transcriptional complex.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=Long;

CC IsoId=O14593-1; Sequences=Displayed;

CC Name=RFX-B-delta5;

CC IsoId=O14593-2; Sequences=VSP_000283, VSP_000284;

CC -!- TISSUE SPECIFICITY: Ubiquitous

CC -!- DOMAIN: The third ankyrin repeat is required for association with the two other RFX subunits; RFX5 and RFXAP.

CC -!- DISEASE: Defects in RFXANK are a cause of bare lymphocyte syndrome type II (BLS II) [MIM:209920]; also known as hereditary MHC class II deficiency or HLA class II-deficient combined immunodeficiency.

CC BLS II is a form of severe combined immunodeficiency disease (SCID) characterized by a profound defect in constitutive and interferon-gamma induced MHC II expression, an absence of cellular and humoral T-cell response to antigen challenge,

CC hypogammaglobulinemia and impaired antibody production. The consequence include extreme susceptibility to viral, bacterial and fungal infections. RFXANK is linked with BLS II complementation group B

CC -!- SIMILARITY: Contains 5 ANK repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

DR EMBL; AF094760; AAC69883.1; -

DR EMBL; AF105427; AAD17972.1; -

DR EMBL; AF105428; AAD17973.1; -

DR EMBL; AF077196; AAD26991.1; -

DR EMBL; AC003110; AAB86654.1; -

DR HSSP; P42773; 1BU9.

DR TRANSFAC; T05441; -

DR TRANSFAC; T05442; -

DR Genew; HGNC:9987; RFXANK.

DR MIM; 603200; -

DR MIM; 209920; -

DR GO; GO:0003677; F:DNA binding; TAS.

DR GO; GO:0003712; F:transcription cofactor activity; TAS.

DR InterPro; IPR002110; ANK.

DR Pfam; PF00023; ank; 4.

DR SMART; SM00248; ANK; 4.

DR PROSITE; PS50088; ANK_REPEAT; 3.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

CC DNA-binding; Transcription regulation; Activator; Nuclear protein; Repeat; ANK repeat; Alternative splicing; Phosphorylation; Disease mutation; SCID; Polymorphism.

CC Repeat 89 118 ANK 1.

CC REPEAT 123 152 ANK 2.

CC REPEAT 156 185 ANK 3.

CC REPEAT 189 218 ANK 4.

CC REPEAT 222 251 ANK 5.

CC REPEAT 63 63

CC VARSPLIC Missing (in isoform RFX-B-delta5).

CC /FTId=VSP_000283.

FT VARSPLIC 91 113 SLSTHQLAAQAGELDQLKEHLKRG -> C (in isoform RFX-B-delta5).

FT FT /FTId=VSP_000284.

FT VARIANT 195 195 L -> P (in BLS II group B).

FT /FTId=VAR_009941.

FT VARIANT 251 251 Q -> E (in dBSNP:1802498).

FT /FTId=VAR_014472.

SQ SEQUENCE 260 AA; 28102 MW; 6280B490F54816D2 CRC64;

Query Match 100.0%; Score 1341; DB 1; Length 260;

Best Local Similarity 100.0%; Pred. No. 1.5e-97;

Matches 260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELTQPAEDLLTQTQTASELGDPEDEGEAAAGSDTVVLSLFFCTPEFVNPEPDASYSS 60

Db 1 MELTQPAEDLLTQTQTASELGDPEDEGEAAAGSDTVVLSLFFCTPEFVNPEPDASYSS 60

Qy 61 PQAGSSLKHSSTLTNRQRGNEVSALPATLDSLSLHQLAAQAGELDQLKEHLKRGDNLVKNP 120

Db 61 PQAGSSLKHSSTLTNRQRGNEVSALPATLDSLSLHQLAAQAGELDQLKEHLKRGDNLVKNP 120

Qy 121 DERGFTPLIWASAFGEIETVRFLEWGDADPHILAKERESALSLASTGGYTDIVGLLLERD 180

Db 121 DERGFTPLIWASAFGEIETVRFLEWGDADPHILAKERESALSLASTGGYTDIVGLLLERD 180

Qy 181 VDINIYDNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKYQQ 240

Db 181 VDINIYDNGGTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKYQQ 240

Qy 241 VIENHILKLFQSNLVPADPE 260

Db 241 VIENHILKLFQSNLVPADPE 260

RESULT 2

RFXK_MOUSE

ID RFXK_MOUSE STANDARD; PRT; 269 AA.

AC Q92205;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE DNA-binding protein RFXANK (Regulatory factor X subunit B) (RFX-B)

DE (Regulatory factor X-associated ankyrin-containing protein) (Ankyrin)

DE Repeat-containing adapter protein Tvl-1).

GN RFXANK OR RFXB OR TVL1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Spleen;

RX MEDLINE=99021383; PubMed=9806546;

RA Masternak K., Barras E., Zufferey M., Conrad B., Corthals G., Abersold R., Sanchez J.-C., Hochstrasser D.F., Mach B., Reith W.;

RT "A gene encoding a novel RFX-associated transactivator is mutated in the majority of MHC class II deficiency patients."

RL Nat. Genet. 20:273-277(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=T-cell;

RX MEDLINE=99262619; PubMed=10329666;

RA Lin J.-H., Makris A., McMahon C., Bear S.E., Patriotis C., Prasad V.R., Brent R., Golemis E.A., Tschlis P.N.;

RT "The ankyrin repeat-containing adaptor protein tvl-1 is a novel substrate and regulator of raf-1."

RL J. Biol. Chem. 274:14706-14715(1999).

RN [3]

RP SEQUENCE FROM N.A. (ISOPFORM SHORT).

RC STRAIN=FVB/N; TISSUE=Salivary gland;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Staphleot M., Soares M.B., Bonaldo M.F., Casavanti T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A.C., Young J.W., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Activates transcription from Class II MHC promoters.
CC Activation requires the activity of the MHC class II
CC transactivator (CIITA). May regulate other genes in the cell (By
CC similarity). Potentiates the activation of RAF-1. RFX binds the X1
CC box of MHC-II promoters.
CC -!- SUBUNIT: Rfx consists of at least 3 different subunits; RFXAP,
CC RFX5 and RFX-B/REFANK. RFX forms cooperative DNA binding complexes
CC with X2BP and CSF/NF-Y. RFX associates with CIITA to form an
CC active transcriptional complex (By similarity). Forms oligomers.
CC Associates with raf-1.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q92205-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q92205-2; Sequence=VSP_000285;
CC TISSUE SPECIFICITY: Expressed primarily in thymus, lung and
CC testis.
CC -!- DOMAIN: Interacts with RAF-1 via its C-terminal ankyrin repeat
CC domain. The same domain also mediates its homodimerization.
CC -!- DOMAIN: The third ankyrin repeat is required for association with
CC the two other RFX subunits; RFX5 and RFXAP (By similarity).
CC -!- PTM: Phosphorylated by RAF-1.
CC -!- SIMILARITY: Contains 3 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF094761; AAC69884.1; -;
CC EMBL; AF123704; AAD24798.1; -;
CC EMBL; BC010971; AAH10971.1; -;
CC HSSP; P80144; 2MYO.
CC TRANSFAC; T05445; -;
CC TRANSFAC; T05446; -;
CC MGD; MGI:1333865; Rfxank.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0007265; P:RAS protein signal transduction; IPI.
CC InterPro; IPR002110; ANK.
CC Pfam; PF00023; ank; 4.
CC SMART; SM00248; ANK; 3.
CC PROSITE; PS50088; ANK REPEAT; 3.
CC PROSITE; PS50297; ANK REP REGION; 1.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein;
KW Repeat; ANK repeat; Alternative splicing; Phosphorylation.
FT REPEAT 88 127 ANK 1.
FT REPEAT 132 161 ANK 2.
FT REPEAT 165 194 ANK 3.
FT REPEAT 198 227 ANK 4.

FT REPEAT 231 260 ANK 5.
FT VARSPLIC 112 121 Missing (in isoform Short).
FT FTId=VSP_000285.
SQ SEQUENCE 269 AA; 29231 MW; 5713F335DC669E87 CRC64;
Query Match 83.0%; Score 1113.5; DB 1; Length 269;
Best Local Similarity 81.9%; Pred. No. 8.7e-80;
Matches 221; Conservative 15; Mismatches 23; Indels 11; Gaps 2;
QY 1 MELTPAEDLIQTOQPASELGDPGPEAGDGSIVLSLFPCTPEPVNPEPDASVSS 60
Db 1 MEPTQVAENLVNQQPPVPLEDPEDRDESPENSIVLSLFPCTPDVNPEDASASS 60
QY 61 PQAGSSLKHSHTLTNRQRNEVSALPATLDSLSIHQAAQGLDQLKHLRK----- 112
Db 61 LQ-GSFLKHSHTLTNRQRNEVSALPATLDSLSIHQAAQGLSOLKDLHKGACACTC 119
QY 113 --GDNLVNPKPDERGFTPLIWAFAFGIEIVRFLLEWADPHILAKRESALSASTGGYT 170
Db 120 LSGNLLINKPDERGFTPLIWAFAFGIEIVRFLLEWADPHILAKRESALSASTGGYT 179
QY 171 DIVGLLLERDVINIDYDNGGTPLLYAVRGNHVKVCVEALLARGADLTTEADSGYTPMDLA 230
Db 180 DIVRLLLRDVINIDYDNGGTPLLYAVRGNHVKVCVEALLARGADLTTEADSGYTPMDLA 239
QY 231 VALGVRKVVQVVIENHILKLFQSNLVPADPE 260
Db 240 VALGVRKVVQVVMESHILRLFQSTLGFVDPE 269
RESULT 3
ANKA HUMAN
ID ANRA HUMAN STANDARD; PRT; 313 AA.
AC Q9H9E1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ankyrin-repeat family A protein 2 (RFXANK-like 2).
GN ANKRA2 OR ANKRA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20547411; PubMed=11095640;
RA Rader K., Orlando R.A., Lou X., Farquhar M.G.;
RT "Characterization of ANKRA, a novel ankyrin repeat protein that
RT interacts with the cytoplasmic domain of megalin.";
RL J. Am. Soc. Nephrol. 11:2167-2178(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Mao Y., Xie Y., Zhou Z., Zhao W., Zhao S., Wang W., Huang Y., Wang S.,
RA Tang R., Chen X., Wu C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:15699-15903(2002).
CC -!- FUNCTION: May facilitate endocytosis by linking megalin to
components of the cytoskeleton or endocytic machinery.
CC -!- SUBUNIT: Interacts directly and specifically with the cytoplasmic
tail of megalin (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and associated with membranes
(BY similarity).
CC -!- SIMILARITY: Contains 3 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF314032; AAK01621.1; -;
CC EMBL; AK022876; BAB14288.1; -;
CC EMBL; AF251051; AKK34941.1; -;
CC EMBL; BC012917; AAL12917.1; -;
CC HSSP; P80144; 2MYO.
CC Genes; HGNC:13208; ANKRA2.
CC MIM; 605787; -;
CC GO; GO:0005829; C:cytosol; IDA.
CC GO; GO:0016020; C:membrane; IDA.
CC GO; GO:0030169; F:low-density lipoprotein binding; IDA.
CC InterPro; IPR002110; ANK.
CC Pfam; PF00023; ank; 3.
CC PRINTS; PR01415; ANKYRIN.
CC SMART; SM00248; ANK; 3.
CC PROSITE; PS50088; ANK_REPEAT; 3.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC Cytoskeleton; ANK repeat; Repeat.
CC REPEAT 181 213 ANK 1.
CC REPEAT 214 246 ANK 2.
CC REPEAT 247 279 ANK 3.
CC SEQUENCE 313 AA; 34272 MW; 31C653B10B4ED6E1 CRC64;
Query Match 46.3%; Score 621.5; DB 1; Length 313;
Best Local Similarity 60.6%; Pred. No. 2.5e-41;
Matches 126; Conservative 26; Mismatches 49; Indels 7; Gaps 2;
QY 51 NPEPDASV-----SSPQAGSSLKHSSTLTNRQGNVEVSALPATLDSLSIHQLAAQGLD 104
DB 104 SPSPGQVRHVYTPSTTKHFSPIKQSTLTNKHGNEVSTTPLLANSLSVHQLAAQGLM 163
QY 105 QLKEHLRGKGNLVNKPDERGFTPLINWASAFGEIETVRFLEWGADPHILAKERESALSIA 164
DB 164 YLATRIEQ-ENVNHTDEEGFTPLMAAAAHGQIAVVEFLQNGADPQLLGKGRESALSIA 222
QY 165 STGGYTDIVGLLERVDVINYDNGGTPLLYAVRGNHVKVCVALLARGADLTTEADSGY 224
DB 223 CSKGYTDIVKMLDCGVDVNEVDWNGGTPLLYAVHGNHVKVCVMLLESADPTIETDSGY 282
QY 225 TPDMLAVALGYRKYVQVQVNIENHILKLFOS 252
DB 283 NSMDLAVALGYRKYVQVQVNIENHILKLFON 310

RESULT 4
ANRA MOUSE
ID ANRA MOUSE STANDARD; PRT; 312 AA.
AC Q99FE2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ankyrin-repeat family A protein 2 (REFANK-like 2).
GN ANKRA2 OR ANKEA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., INTERACTION WITH MEGALIN, AND SUBCELLULAR
RP LOCATION.
RX MEDLINE=20547411; PubMed=11095640;
RA Rader K., Orlando R.A., Lou X., Farquhar M.G.;
RT "Characterization of ANKRA, a novel ankyrin repeat protein that
interacts with the cytoplasmic domain of megalin";
RL J. Am. Soc. Nephrol. 11:2167-2178(2000).
CC -!- FUNCTION: May facilitate endocytosis by linking megalin to
components of the cytoskeleton or endocytic machinery.
CC -!- SUBUNIT: Interacts directly and specifically with the cytoplasmic
tail of megalin.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and associated with membranes.
CC -!- SIMILARITY: Contains 3 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF314031; AAK01620.1; -;
CC HSSP; P80144; 2MYO.
CC MGD; MGI:1915808; Ankra2.
CC GO; GO:0005624; C:membrane fraction; IDA.
CC GO; GO:0005515; F:protein binding; IPT.
CC InterPro; IPR002110; ANK.
CC Pfam; PF00023; ank; 3.
CC PRINTS; PR01415; ANKYRIN.
CC SMART; SM00248; ANK; 3.
CC PROSITE; PS50088; ANK_REPEAT; 3.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC Cytoskeleton; ANK repeat; Repeat.
CC REPEAT 180 212 ANK 1.
CC REPEAT 213 245 ANK 2.
CC REPEAT 246 278 ANK 3.
CC SEQUENCE 312 AA; 34062 MW; 3B52B1415B7A5AFA CRC64;
Query Match 45.7%; Score 612.5; DB 1; Length 312;
Best Local Similarity 60.1%; Pred. No. 1.3e-40;
Matches 125; Conservative 25; Mismatches 51; Indels 7; Gaps 2;
QY 51 NPEPDASV-----SSPQAGSSLKHSSTLTNRQGNVEVSALPATLDSLSIHQLAAQGLD 104
DB 103 SPSPGQVRHVYTPSTTKHFSPIKQSTLTNKHGNEVSTTPLLANSLSAHQLAAQGLM 162
QY 105 QLKEHLRGKGNLVNKPDERGFTPLINWASAFGEIETVRFLEWGADPHILAKERESALSIA 164
DB 163 YLATRIEQ-ENVNHTDEEGFTPLMAAAAHGQIAVVEFLQNGADPQLLGKGRESALSIA 221
QY 165 STGGYTDIVGLLERVDVINYDNGGTPLLYAVRGNHVKVCVALLARGADLTTEADSGY 224
DB 222 CSKGYTDIVKMLDCGVDVNEVDWNGGTPLLYAVHGNHVKVCVMLLENADPTIETDSGY 281
QY 225 TPDMLAVALGYRKYVQVQVNIENHILKLFOS 252
DB 282 NSMDLAVALGYRKYVQVQVNIENHILKLFON 309

RESULT 5
 HTM1_HUMAN STANDARD; PRT; 1267 AA.
 AC Q9H9B1; Q8TCN7; Q9GF53; Q96J11; Q96KH4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Histone-lysine N-methyltransferase, H3 lysine-9 specific 5
 DE (EC 2.1.1.43) (Histone H3-K9 methyltransferase 5) (H3-K9-HMTase 5)
 DE (Euchromatic histone methyltransferase 1) (Eu-HMTase1) (G9a-like
 DE protein 1) (GLP1).
 GN EUHMTASE1 OR KIAA1876.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), ENZYMIC ACTIVITY, AND IDENTIFICATION
 RP IN COMPLEX WITH E2F6; TFDPI; MAX; MGA; BAT8; CBX3; RING1; RNF2; MBLR;
 RP L3MBTL2 AND YAF2.
 RC TISSUE=Cervical carcinoma;
 RX MEDLINE=2199559; PubMed=12004135;
 RA Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.;
 RT "A complex with chromatin modifiers that occupies E2F- and
 RT Myc-responsive genes in G0 cells.";
 RL Science 296:1132-1136(2002).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Teratocarcinoma;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakanatsu A., Nakamura Y., Nagahara K., Masuo Y., Oshima A.;
 RT "NEDD human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 351-1267 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RA Tsuritani K., Ukai Y., Yajima Y., Amemiya C., Yoshimoto M.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 351-1267 FROM N.A. (ISOFORM 3).
 RX MEDLINE=22158633; PubMed=12168954;
 RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
 RT "Construction of expression-ready cDNA clones for KIAA genes: manual
 RT curation of 330 KIAA cDNA clones.";
 RL DNA Res. 9:99-106(2002).
 RN [5]
 RP SEQUENCE OF 1150-1267 FROM N.A. (ISOFORM 3), AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=21245130; PubMed=11347906;
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 8:85-95(2001).
 RN [6]
 RP SEQUENCE OF 552-1267 FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshikiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Maman A., Rodriguez S., Sanchez A.,
 RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RP SEQUENCE OF 1056-1267 FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymph node;
 RA Ansoorge W., Wirkner U., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION. Histone methyltransferase. Methylates Lys-9 of histone
 CC H3 (in vitro). H3 Lys-9 methylation represents a specific tag for
 CC epigenetic transcriptional repression by recruiting HPI proteins
 CC to methylated histones. Probably targeted to histone H3 by
 CC different DNA-binding proteins like E2F6, MGA, MAX and/or DP1.
 CC During G0 phase, it probably contributes to silencing of MYC- and
 CC E2F-responsive genes, suggesting a role in G0/G1 transition in
 CC cell cycle.
 CC -1- CATALYTIC ACTIVITY. S-adenosyl-L-methionine + histone L-lysine =
 CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
 CC -1- SUBUNIT. Part of the E2F6.com-1 complex in G0 phase composed of
 CC E2F6, MGA, MAX, TFDPI, CBX3, BAT8, EUHMTASE1, RING1, RNF2, MBLR,
 CC L3MBTL2 and YAF2.
 CC -1- SUBCELLULAR LOCATION. Nuclear; associates with euchromatic
 CC regions.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;
 CC IsoId=Q9H9B1-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9H9B1-2; Sequence=VSP_002222, VSP_002223;
 CC Name=3;
 CC IsoId=Q9H9B1-3; Sequence=VSP_002224, VSP_002225;
 CC -1- TISSUE SPECIFICITY. Widely expressed.
 CC -1- SIMILARITY. Belongs to the histone-lysine methyltransferase
 CC family.
 CC -1- SIMILARITY. Contains 8 ANK repeats.
 CC -1- SIMILARITY. Contains 1 pre-SET domain.
 CC -1- SIMILARITY. Contains 1 SET domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AY083210; AA09024.1; -
 CC EMBL; AK022941; BAB14321.1; -
 CC EMBL; AB028932; BAB56104.1; -
 CC EMBL; AB058779; BAB47505.2; -
 CC EMBL; BC011608; AAL11608.2; -
 CC EMBL; AL713772; CAD28534.1; ALT_INIT.
 CC MIM; 607001; -
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR007728; Pre-SET.
 CC InterPro; IPR001214; SET.
 CC InterPro; IPR003606; Zn2-binding.
 CC Pfam; PF00023; ank; 7.
 CC Pfam; PF05033; Pre-SET; 1.
 CC Pfam; PF00856; SET; 1.
 CC PRINTS; PR01415; ANKYRIN.
 CC SMART; SM00248; ANK; 7.
 CC SMART; SM00468; PreSET; 1.
 CC SMART; SM00317; SET; 1.

```

DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50867; PRE_SET; 1.
DR PROSITE; PS50280; SET; 1.
KW Transferrase; Methyltransferase; Chromatin regulator; Nuclear protein;
KW ANK repeat; Repeat; Alternative splicing.
FT REPEAT 706 735 ANK 1.
FT REPEAT 741 770 ANK 2.
FT REPEAT 774 803 ANK 3.
FT REPEAT 807 837 ANK 4.
FT REPEAT 841 870 ANK 5.
FT REPEAT 874 903 ANK 6.
FT REPEAT 907 936 ANK 7.
FT REPEAT 940 973 ANK 8.
FT DOMAIN 1029 1092 PRE-SET.
FT DOMAIN 1094 1216 SET.
FT DOMAIN 375 378 POLY-GLU.
FT DOMAIN 411 418 POLY-ARG.
FT DOMAIN 1261 1264 POLY-ALA.
FT VARSPPLIC 805 825 DAEGSTCLHLAAKKGHVEVQ -> IQKTSKVVTESQETQR
FT VARSPPLIC 826 1267 Missing (in isoform 2).
FT VARSPPLIC 826 1267 Missing (in isoform 2).
FT VARSPPLIC 1150 1153 DGEV -> ISSA (in isoform 3).
FT VARSPPLIC 1154 1267 Missing (in isoform 3).
FT VARSPPLIC 1154 1267 Missing (in isoform 3).
FT CONFLICT 530 530 G -> E (IN REF. 3 AND 4).
SQ SEQUENCE 1267 AA; 138181 MW; 78FD418FD404F475 CRC64;

Query Match 15.4%; Score 206.5; DB 1; Length 1267;
Best Local Similarity 29.6%; Pred. No. 3.7e-08;
Matches 55; Conservative 33; Mismatches 81; Indels 17; Gaps 3;

QY 62 QAGSSL-----KSTILTNRQNEVSLP-----ATLD-----SLSHQLAAGELD 104
DQ 763 QAGANIDTCSQDQRTPLMEAAENHLEAVKYLKAGALVDPKDAEGSTCLHLAAKKGHVE 822
QY 105 QLKEHLKGDNLNKPDERGFTPLWASAFGETETVFLLEMGADPHILAKRESALSIA 164
DQ 823 VVGILSNGQMDVNCQDGGTWMWATEYKHVYKLLSKGSDINRDNEENICLHWA 882
QY 165 STGYTDIVGLLERVDVINIYDNGGTPLLYAVRGVNHVKCVALLARGADLTTEADSGY 224
DQ 883 AFSGCVDIABILLAAKCDLHAVNHGDSPLHIAARENRYDCVVLFLSRSDSVTLKNKEG 942
QY 225 TPDMLA 230
DQ 943 TPLQCA 948

RESULT 6
ID ANK2_HUMAN STANDARD; PRT; 3924 AA.
AC Q01484; Q01485;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN ANK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Brain stem;
RA Chan W., Kordeli E., Bennett V.;
RX MEDLINE=91302466; PubMed=1830053;
RT "Isolation and characterization of cDNAs encoding human brain
RT ankyrins reveal a family of alternatively spliced genes.";
RL J. Cell Biol. 114:241-253(1991).

[2]
RN REVISIONS.
RP Carpenter S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordeli E., Bennett V.;
RT "440-kD ankyrinB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons.";
RL J. Cell Biol. 123:1463-1473(1993).
RN [4]
RP SEQUENCE OF 463-495 FROM N.A.
RX MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankyrin gene.";
RL Genomics 10:858-866(1991).
CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC elements. Also bind to cytoskeletal proteins.
CC -!- ALTERNATIVE PRODUCTS.
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoID=Q01484-1; Sequence=VSP_000267; VSP_000268;
CC Name=2;
CC IsoID=Q01484-2; Sequence=VSP_000267; VSP_000268;
CC Name=3;
CC IsoID=Q01484-3; Sequence=VSP_000268;
CC -!- TISSUE SPECIFICITY: Plasma membrane of neurons as well as glial
CC cells throughout the brain.
CC -!- PTM: Phosphorylated at multiple sites by different protein kinases
CC and each phosphorylation event regulates the protein's structure
CC and function (Potential).
CC -!- SIMILARITY: Contains 23 ANK repeats.
CC -!- SIMILARITY: Contains 1 death domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56957; CAA40278.1; -
DR EMBL; X56958; CAA40279.2; -
DR EMBL; Z26634; CAB42644.1; -
DR EMBL; M37123; AAA62828.1; -
DR PIR; S37431; S37431.
DR HSP; P42771; IDC2.
DR Genew; HGNC:493; ANK2.
DR MIM; 106410; -
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000488; Death.
DR InterPro; IPR000906; ZUS.
DR Pfam; PF00023; ank; 24.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00791; ZUS; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 22.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00218; ZUS; 1.
DR PROSITE; PS50088; ANK_REPEAT; 20.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation.
FT REPEAT 63 92 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 129 158 ANK 3.
FT REPEAT 162 191 ANK 4.

```

```

FT REPEAT 193 220 ANK 5.
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A.
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A.
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH.
FT VARSPPLIC 1039
Q -> QFLGKLHPTAPPLNEGESLSVRLQGPPTK
(in isoform 2).
/FTid=VSP 000267.
Missing (in isoform 2 and isoform 3).
FT VARSPPLIC 1444 3528
Q -> PE (IN REF. 4).
/FTid=VSP 000268.
FT CONFLICT 475 476
FT CONFLICT 971 971 I -> S (IN REF. 1).
FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).
FT CONFLICT 3586 3586 I -> Y (IN REF. 1).
SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;

Query Match 15.1%; Score 202.5; DB 1; Length 3924;
Best Local Similarity 29.3%; Pred. No. 3.4e-07;
Matches 56; Conservative 43; Mismatches 81; Indels 11; Gaps 4;

QY 55 DASVSPQAGSSIKHSTTLTNRGRNEVSALPATLDSLSIHLAAQAGELDLQKEHLRKGD 114
Db 5 DAAQKS-DSGERENGSSQRRKPKKSDSNA-----SFLRAARAGNLDKVEYLKGGI 55
QY 115 NLVKNKDEGFTEPLIWAFAGETVTRFLEMGADPHILAKRESALSASTGGYDPIVG 174
Db 56 D-INTCNGNLNALHLAAKEGHVGLVQELLGRGSSVDSATKKGNTALHTIASLAGQAEVVK 114
QY 175 LLERDVIDNIYDWNGTPLLAVRGNHVKCYEALLARGADLTTEADSGYTPMDLVALIG 234
Db 115 VLVKEGANIASQNGFTPLVYAAQENHDVVKYLLENGANOSTATEDGFTFLAVALQQG 174
QY 235 YRK-VQOVLEN 244
Db 175 HNOQAVILLEN 185

RESULT 7
ANK3 HUMAN STANDARD; PRT; 4377 AA.
ID ANK3 HUMAN
AC Q12955;
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```


RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Fulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP ALTERNATIVE SPLICING (ISOFORM 2).
RX MEDLINE=21564388; PubMed=11707778;
RA Brown S.E., Campbell R.D., Sanderson C.M.;
RT "Novel NG36/G9a gene products encoded within the human and mouse MHC
RT class III regions.";
RL Mamm. Genome 12:916-924(2001).
RN [5]
RP ENZYMACTIC ACTIVITY, SUBCELLULAR LOCATION, AND MUTAGENESIS OF ARG-1162.
RX MEDLINE=21326082; PubMed=11316813;
RA Tachibana M., Sugimoto K., Fukushima T., Shinkai Y.;
RT "Set domain-containing protein, G9a, is a novel lysine-preferring
RT mammalian histone methyltransferase with hyperactivity and specific
RT selectivity to lysines 9 and 27 of histone H3.";
RL J. Biol. Chem. 276:25309-25317(2001).
CC -!- FUNCTION: Histone methyltransferase. Preferentially methylates
CC Lys-9 of histone H3 and Lys-27 of histone H3 (in vitro). H3 Lys-9
CC methylation represents a specific tag for epigenetic
CC transcriptional repression by recruiting H3 proteins to
CC methylated histones. Also methylates histone H1.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =
CC S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.
CC -!- SUBUNIT: Part of the E2F6-com-1 complex in G0 phase composed of
CC E2F6, MGA, MAX, TFDPI, CSX3, BAT8, EUMTASE1, RING1, RNF2, MBLR,
CC LIMB12 and YAP2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear; almost excluded from nucleoli.
CC Associates with euchromatic regions. Does not associate with
CC heterochromatin.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=G9a-L;
CC IsoId=Q9Z148-1; Sequence=Displayed;
CC Name=2; Synonyms=G9a-S;
CC IsoId=Q9Z148-2; Sequence=VSP_002214, VSP_002215, VSP_002216;
CC -!- MISCELLANEOUS: G9a deficient mice show a higher level of histone
CC H3 with acetylated Lys-9 and/or methylated Lys-4, display severe
CC developmental defects and die within E9.5-E12.5 stages.
CC -!- SIMILARITY: Belongs to the histone-lysine methyltransferase
CC family. Suvar3-9 subfamily.
CC -!- SIMILARITY: Contains 7 ANK repeats.
CC -!- SIMILARITY: Contains 1 pre-SET domain.
CC -!- SIMILARITY: Contains 1 SET domain.
CC -!- SIMILARITY: Contains 1 post-SET domain.
CC -!- CAUTION: NG36 and G9a were originally thought to derive from two
CC separate genes.
CC -!- CAUTION: Ref.1 sequences differ from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF109906; AAC84164.1; ALT_SEQ.
CC EMBL; AF109906; AAC84165.1; ALT_SEQ.

DR EMBL; AB077209; BAC05482.1; -;
DR EMBL; AB077210; BAC05483.1; -;
DR EMBL; BC025539; AAH25539.1; ALT_INIT.
DR MGD; MGI:2148922; Bat8.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR003616; PostSET.
DR InterPro; IPR007728; Pre-SET.
DR InterPro; IPR001214; SET.
DR InterPro; IPR003606; Zn2-binding.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF05033; Pre-SET; 1.
DR Pfam; PF00856; SET; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00468; PreSET; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS50088; ANK_REPEAT; 5.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50868; POST_SET; FALSE_NEG.
DR PROSITE; PS50867; PRE_SET; 1.
DR PROSITE; PS50280; SET; 1.
KW Transferase; Methyltransferase; Chromatin regulator; Nuclear protein;
KW ANK repeat; Repeat; Alternative splicing.
FT DOMAIN 352 379 POLY-GLU.
FT REPEAT 702 731 ANK 1.
FT REPEAT 737 766 ANK 2.
FT REPEAT 770 799 ANK 3.
FT REPEAT 803 833 ANK 4.
FT REPEAT 837 866 ANK 5.
FT REPEAT 870 899 ANK 6.
FT REPEAT 903 932 ANK 7.
FT DOMAIN 1025 1088 PRE-SET.
FT DOMAIN 1090 1212 SET.
FT DOMAIN 1217 1233 POST-SET.
FT VARSPIC 1 57 Missing (in isoform 2).
FT VARSPIC 58 71 Missing (in isoform 2).
FT VARSPIC 426 459 Missing (in isoform 2).
FT VARSPIC 1162 1162 Missing (in isoform 2).
FT MUTAGEN 1162 1162 R->H; STRONGLY REDUCES HISTONE
FT MUTAGEN 1165 1168 METHYLTRANSFERASE ACTIVITY.
FT MUTAGEN 1165 1168 MISSING; ABOLISHES HISTONE
FT MUTAGEN 1165 1168 METHYLTRANSFERASE ACTIVITY AND SUBSEQUENT
FT MUTAGEN 1165 1168 REPRESSION.
SQ SEQUENCE 1263 AA; 138038 MW; 74DBFF9A36769589 CRC64;
Query Match 14.0%; Score 188; DB 1; Length 1263;
Best Local Similarity 34.5%; Pred. No. 1e-06;
Matches 48; Conservative 19; Mismatches 72; Indels 0; Gaps 0;
QY 91 SLISHQLAAGGELDQLKEHLKGNLVNKPDERGFTPLIWAFAFGEIETVRFLEWGADP 150
DB 805 STCLHAAKIGNLEWVSLLSSTGQVDVNAQDSGWTPIIWAHEKHIDIVRMLLTGADV 864
QY 151 HILAKERESALSASTGGYTDIVGLLERDVIDINIDMNGGTPLLYAVRGNHVKCVFALL 210
DB 865 TLTDNEENICLHWASFTGSAALAEVLLNAQCDLHAVNVHGDTPHIAARESVDCLVFL 924
QY 211 ARGADLTTEADSGYTPMDL 229
DB 925 SRGANPELNKSGDTAWDL 943
RESULT 10
ID GASZ_MOUSE STANDARD; PRT; 475 AA.
AC Q9VD46; Q9YKQ7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Germ cell-specific ankyrin, SAM and basic leucine zipper domain

CC Binds to the N-terminus of telomeric TRF1 via the ANK repeats.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic, associated with the Golgi and
 CC with juxtanuclear SLCA4/GLUT4-vesicles. A minor proportion is
 CC also found at nuclear pore complexes and around the pericentriolar
 CC matrix of mitotic centrosomes. During interphase, a small fraction
 CC of TNKS is found in the nucleus, associated with TRF1.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Name=1;
 CC IsoId=095271-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=095271-2; Sequence=VSP_004538, VSP_004539;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Ubiquitous; highest levels in testis.
 CC -1- PTM: Upon insulin-stimulation, phosphorylated on serine residues
 CC by MAPK kinases.
 CC -1- PM: ADP-ribosylated (-auto).
 CC -1- SIMILARITY: Belongs to the PARP family.
 CC -1- SIMILARITY: Contains 15 ANK repeats.
 CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AF082556; AAC79841.1; -;
 DR EMBL; AF082557; AAC79842.1; -;
 DR EMBL; AF082558; AAC79843.1; -;
 DR EMBL; AF082559; AAC79844.1; -;
 DR HSPSP; Q00420; 1AWC.
 DR Genew; HGNC:11941; TNKS.
 DR MIM; 603303; -;
 DR GO; GO:0000781; C:chromosome, telomeric region; IDA.
 DR GO; GO:0003950; F:NAD ADP-ribosyltransferase activity; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001660; SAM.
 DR Pfam; PF00023; ank; 19.
 DR Pfam; PF00536; SAM; 1.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SM00248; ANK; 17.
 DR SMART; SM00454; SAM; 1.
 DR PROSITE; PSS0088; ANK_REPEAT; 15.
 DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PSS0105; SAM_DOMAIN; 1.
 KW Transferase; Glycosyltransferase; NAD; Golgi stack; Telomere;
 KW Nuclear protein; Repeat; ANK repeat; ADP-ribosylation;
 KW Phosphorylation; Alternative splicing.
 FT REPEAT 215 247 ANK 1.
 FT REPEAT 248 280 ANK 2.
 FT REPEAT 281 313 ANK 3.
 FT REPEAT 368 400 ANK 4.
 FT REPEAT 401 433 ANK 5.
 FT REPEAT 434 466 ANK 6.
 FT REPEAT 521 556 ANK 7.
 FT REPEAT 557 589 ANK 8.
 FT REPEAT 590 622 ANK 9.
 FT REPEAT 683 715 ANK 10.
 FT REPEAT 716 748 ANK 11.
 FT REPEAT 749 781 ANK 12.
 FT REPEAT 836 868 ANK 13.
 FT REPEAT 869 901 ANK 14.
 FT REPEAT 902 934 ANK 15.
 FT REPEAT 1030 1089 SAM.
 FT DOMAIN 1176 1327 PARP.
 FT DOMAIN 9 14 POLY-HIS.
 FT DOMAIN 27 34 POLY-PRO.
 FT DOMAIN 128 134 POLY-SER.

FT	DOMAIN	137	145	POLY-SER.
FT	VARSPLIC	641	643	EST -> GHS (in isoform 2).
FT				/FTId=VSP_004538.
FT	VARSPLIC	644	1327	Missing (in isoform 2).
FT				/FTId=VSP_004539.
FT	MUTAGEN	1184	1184	H->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
FT				WITH A-1291.
FT	MUTAGEN	1291	1291	E->A: LOSS OF ACTIVITY; WHEN ASSOCIATED
FT				WITH A-1184.
SQ	SEQUENCE	1327	AA; 142010	MW; E14DE985C710B957 CRC64;

Query Match 13.9%; Score 187; DB 1; Length 1327;
 Best local similarity 24.3%; Pred. No. 1.3e-06;
 Matches 72; Conservative 36; Mismatches 112; Indels 76; Gaps 9;

QY	18	ASELGPEPEGGEADGSDTV-VLSFPCTPEPVNPEPDASVSSPGAGSLKHSTLTNR	76
Db	137	SSSSSPSPGSLAESPEAGVSSTAPLPGAGNP----	GTGVPASGALRE---LLEA 189
QY	77	QRGNEVSALPATLDSLSI-----	HQLAAQGLDQJKEHLRKGDNIWNKXPD 123
Db	190	CRNGDVSRVKRLVDAANNAKDMAGKSGPLHFAAGFGRKDVVHELLQMGAN-VHARDG	248
QY	124	GFTPLWAGAFGEIEVRFLL-----	EW-----GADP 150
Db	249	GLIPLNACSGFHAEVVSLLCQAGADPNARDNNNYTFLHEAAIKGKIDVCIVLLQHGADP	308
QY	151	HILAKEREGALSILAS-----	TGGY-----TDIVGLLLERDVINIYDWN 189
Db	309	NIRNTDGSALDLADPSAKAVLTGEYKDELLEAARSNGEEKLMALLPLNVNCHASDGR	368
QY	190	GGTPLLAVRGNHVKCEALLARGADLTTEADSGYTPMDLAVGALGYRKKVQOVIENH	245
Db	369	KSTPLHLAGYNNRVIRVQLLQHGADVHAKDGLVPLHNACSYGHVEVTELLKH	424

RESULT 12

ID	ANK1	HUMAN	STANDARD;	PRT;	1880	AA.
AC	P16157;					
DT	01-APR-1990	(Rel. 14, Created)				
DT	01-APR-1990	(Rel. 14, Last sequence update)				
DT	15-MAR-2004	(Rel. 43, Last annotation update)				
DE	Ankyrin 1 (Erythrocyte ankyrin)	(Ankyrin R).				
GN	ANK1	OR ANK.				
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.					
OX	NCHI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND VARIANTS.					
RC	TISSUE=Hematopoietic;					
RX	MEDLINE=90158830; PubMed=2137557;					
RA	Lux S.E., John K.M., Bennett V.;					
RT	"Analysis of cDNA for human erythrocyte ankyrin indicates a repeated					
RT	structure with homology to tissue-differentiation and cell-cycle					
RT	control proteins."					
RL	Nature 344:36-42(1990).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=90175370; PubMed=1689849;					
RA	Lambert S., Yu H., Prchal J.T., Lawler J., Ruff P., Speicher D.,					
RA	Cheung M.C., Kan Y.W., Palek J.;					
RT	"cDNA sequence for human erythrocyte ankyrin."					
RL	Proc. Natl. Acad. Sci. U.S.A. 87:1730-1734(1990).					
RN	[3]					
RP	VARIANT HS ILE-462.					
RX	MEDLINE=96225450; PubMed=8640229;					
RA	Eber S.W., Gonzalez J.M., Lux M.L., Scarpa A.L., Tse W.T.,					
RA	Dornwell M., Herbers J., Kugler W., Oezcan R., Pekrun A.,					
RA	Gallagher P.G., Schroeter W., Forget B.G., Lux S.E.;					
RT	"Ankyrin-1 mutations are a major cause of dominant and recessive					
RT	hereditary spherocytosis."					

RL Nat. Genet. 13:214-218 (1996).

CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal elements; bind to the erythrocyte membrane protein band 4.2, to Na-K ATPase, to the lymphocyte membrane protein GP85, and to the cytoskeletal proteins fodrin, tubulin, vimentin and desmin.

CC Erythrocyte ankyrins also link spectrin (beta chain) to the cytoplasmic domain of the erythrocytes anion exchange protein; they retain most or all of these binding functions.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF ERYTHROCYTIC PLASMA MEMBRANE.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=3;

CC Comment=Additional isoforms seem to exist;

CC Name=1; Synonyms=2.1;

CC IsoId=PI6157-1; Sequence=displayed;

CC Name=2; Synonyms=2.2;

CC IsoId=PI6157-2; Sequence=VSP_000264, VSP_000265;

CC Name=3;

CC IsoId=PI6157-3; Sequence=VSP_000266;

CC -!- PTM: Regulated by phosphorylation.

CC -!- PTM: Palmitoylated.

CC -!- DISEASE: Defects in ANK1 are a cause of hereditary spherocytosis (HS) [MIM:182900]. Inheritance can be autosomal dominant or recessive.

CC -!- SIMILARITY: Contains 23 ANK repeats.

CC -!- SIMILARITY: Contains 1 death domain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X16609; CAA34610.1; -;

DR EMBL; M28880; AAS1732.1; -;

DR PIR; A35049; A35049.

DR PIR; S08275; SUHUK.

DR HSSP; Q00420; LAWC.

DR Genew; HGNC:492; ANK1.

DR MIM; 182900; -;

DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000488; Death.

DR InterPro; IPR000906; ZU5.

DR Pfam; PF00023; ank; 24.

DR Pfam; PF00531; death; 1.

DR Pfam; PF00791; ZU5; 1.

DR PRINTS; PR01415; ANKYRIN.

DR SMART; SM00248; ANK; 21.

DR SMART; SM00218; ZU5; 1.

DR PROSITE; PS50088; ANK_REPEAT; 20.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS50017; DEATH_DOMAIN; 1.

KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;

KW Phosphorylation; Lipoprotein; Disease mutation; Elliptocytosis;

KW Polymorphism.

KW INIT MET 0 0

FT DOMAIN 1 826 89 kDa DOMAIN

FT DOMAIN 827 1381 (ANION EXCHANGE PROTEIN BINDING DOMAIN).

FT DOMAIN 1382 1880 62 kDa DOMAIN (SPECTRIN BINDING DOMAIN).

FT 55 kDa REGULATORY DOMAIN

FT (REGULATES THE BINDING OF ANKYRIN TO

FT SPECTRIN AND THE BAND 3 PROTEIN).

FT REPEAT 43 72 ANK 1.

FT REPEAT 76 105 ANK 2.

FT REPEAT 109 138 ANK 3.

FT REPEAT 142 171 ANK 4.

FT REPEAT 173 200 ANK 5.

FT REPEAT 204 233 ANK 6.

FT REPEAT 237 266 ANK 7.

FT REPEAT 270 299 ANK 8.

FT REPEAT 303 332 ANK 9.

FT REPEAT 336 365 ANK 10.

FT REPEAT 369 398 ANK 11.

FT REPEAT 402 431 ANK 12.

FT REPEAT 435 464 ANK 13.

FT REPEAT 468 497 ANK 14.

FT REPEAT 501 530 ANK 15.

FT REPEAT 534 563 ANK 16.

FT REPEAT 567 596 ANK 17.

FT REPEAT 600 629 ANK 18.

FT REPEAT 633 662 ANK 19.

FT REPEAT 666 695 ANK 20.

FT REPEAT 699 728 ANK 21.

FT REPEAT 732 761 ANK 22.

FT REPEAT 765 794 ANK 23.

FT DOMAIN 1402 1486 DEATH.

FT VARSPPLIC 1512 1873 Missing (in isoform 2).

FT VARSPPLIC 1874 1874 H -> D (in isoform 2).

FT VARSPPLIC 1849 1880 TVEGPLEDPSELEVDIDYFMKHSKDHSTPNP -> ELRGS

FT 3). GLQPLIEGRKGAQIVKRASLRKQKQ (in isoform

FT /FTId=VSP_000266.

FT R -> T.

FT /FTId=VAR_000595.

FT V -> I (in HS).

FT /FTId=VAR_000596.

FT R -> H (in Brueggen).

FT /FTId=VAR_000597.

FT V -> A.

FT /FTId=VAR_000598.

FT D -> E.

FT /FTId=VAR_000599.

FT E -> D.

FT /FTId=VAR_000601.

FT S -> T.

FT /FTId=VAR_000600.

FT D -> N (in Dueseldorf).

FT /FTId=VAR_000602.

FT R -> D.

FT /FTId=VAR_000603.

FT A -> S (IN REF. 2).

FT CONFLICT 229 229

FT CONFLICT 1545 1545 V -> I (IN REF. 2).

SQ SEQUENCE 1880 AA; 206145 MW; 1C5F5E7EFD1CD428 CRC64;

Query Match 13.7%; Score 184; DB 1; Length 1880;

Best Local Similarity 23.1%; Pred. No. 3.6e-06;

Matches 54; Conservative 37; Mismatches 69; Indels 74; Gaps 3;

QY 74 TNQRQNEVSALPATLDSLSIHOLAAQGLDOLKEHLKGDNLVKNPDERGFTPLIWSA 133

DB 72 TTTKGN-----TALHIALAGDEVVRELNVYGAN-VNAQSQKGTPLVMAAQ 119

QY 134 FGEIETVRLLEWGPADPHILAKERESALSLASTGGYTDIVG----- 174

DB 120 ENHLEWVKFLLENGANQVATEDGTPLAVALQOQHENVVAHLINVTGKVKRLPALHIA 179

QY 175 -----LLLRVDVDINIDWNGG 191

DB 180 ARNDTTRTAAVLLNDPNPDVLSKTGFTPLHIAHYENLNVAQLLNRGASVNFPTPQNGI 239

QY 192 TPLLYAVRGNHVKVCEALLARGADLTTEADSGYTPMDLAVALGYRKVQGVINRH 245

DB 240 TPLHIAARRGNVIMVRLDLLDRGAQIETKTKDELTPHCAARNGHVRISEILDH 293

RESULT 13

ANK1 MOUSE

ID ANK1 MOUSE STANDARD; PRT; 1862 AA.

AC Q02357;

DT 01-NOV-1995 (Rel. 32, Created)

01-NOV-1995 (Rel. 32, Last sequence update)
 16-OCT-2001 (Rel. 40, Last annotation update)
 Ankyrin 1 (Erythrocyte ankyrin).
 ANK1 OR ANK-1.
 Mus musculus (Mouse).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 NCBI_TaxID=10090;
 SEQUENCE FROM N.A.
 TISSUE=Erythrocyte;
 MEDLINE=92345717; PubMed=1386265;
 White R.A., Birkenmeier C.S., Peters L.L., Barker J.E., Lux S.E.;
 "Murine erythrocyte ankyrin cDNA: highly conserved regions of the
 regulatory domain";
 Mamm. Genome 3:281-285(1992).
 CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
 elements; bind to the erythrocyte membrane protein band 4.2, to the
 Na-K ATPase, to the lymphocyte membrane protein gp85, and to the
 cytoskeletal proteins fodrin, tubulin, vimentin and desmin.
 CC Erythrocyte ankyrins also link spectrin (beta chain) to the
 cytoplasmic domain of the erythrocytes anion exchange protein;
 CC they retain most or all of these binding functions.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF ERYTHROCYTIC
 PLASMA MEMBRANE.
 CC -!- PTM: Regulated by phosphorylation (by similarity).
 CC -!- PTM: Acylated by palmitic acid group(s) (By similarity).
 CC -!- SIMILARITY: Contains 23 ANK repeats.
 CC -!- SIMILARITY: Contains 1 death domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M84756; AAA37236.1; -.
 CC PIR; I49502; I49502.
 CC HSPP; Q00420; IAWC.
 CC MGD; MGI.88024; Ankl.
 CC InterPro; IPR002110; ANK.
 CC InterPro; IPR000488; Death.
 CC InterPro; IPR000906; ZU5.
 CC Pfam; PF00023; ank; 24.
 CC Pfam; PF00531; death; 1.
 CC Pfam; PF00791; ZU5; 1.
 CC PRINTS; PR01415; ANKYRIN.
 CC SMART; SM00248; ANK; 22.
 CC SMART; SM00005; DEATH; 1.
 CC SMART; SM00218; ZU5; 1.
 CC PROSITE; PS50088; ANK REPEAT; 20.
 CC PROSITE; PS50297; ANK REP REGION; 1.
 CC PROSITE; PS50017; DEATH DOMAIN; 1.
 KW Cytoskeleton; Repeat; ANK repeat; Phosphorylation; Lipoprotein.
 FT DOMAIN 1 827
 FT 89 kDa DOMAIN (ANION EXCHANGE PROTEIN
 BINDING DOMAIN).
 FT
 FT DOMAIN 828 1386
 FT 62 kDa DOMAIN (SPECTRIN BINDING
 DOMAIN).
 FT
 FT DOMAIN 1387 1862
 FT 55 kDa REGULATORY DOMAIN (REGULATES
 THE BINDING OF ANKYRIN TO SPECTRIN
 AND THE BAND 3 PROTEIN).
 FT
 FT REPEAT 40 69
 FT REPEAT 73 102
 FT REPEAT 106 135
 FT REPEAT 139 168
 FT REPEAT 170 197
 FT REPEAT 201 230
 FT REPEAT 234 263
 FT REPEAT 267 296
 FT REPEAT 300 329
 FT REPEAT 333 362

FT REPEAT 366 395 ANK 11.
 FT REPEAT 399 428 ANK 12.
 FT REPEAT 432 461 ANK 13.
 FT REPEAT 465 494 ANK 14.
 FT REPEAT 498 527 ANK 15.
 FT REPEAT 531 560 ANK 16.
 FT REPEAT 564 593 ANK 17.
 FT REPEAT 597 626 ANK 18.
 FT REPEAT 630 659 ANK 19.
 FT REPEAT 663 692 ANK 20.
 FT REPEAT 696 725 ANK 21.
 FT REPEAT 729 758 ANK 22.
 FT REPEAT 762 791 ANK 23.
 FT DOMAIN 1399 1483 DEATH.
 SQ SEQUENCE 1862 AA; 204242 MW; A56B85B5B29001E5 CRC64;
 Query Match 13.6%; Score 183; DB 1; Length 1862;
 Best Local Similarity 23.1%; Pred. No. 4.2e-06;
 Matches 54; Conservative 36; Mismatches 70; Indels 74; Gaps 3;
 QY 74 TNEQRGNEVSALPATLDSLSIHQAQGGELDQLKEHLKRGDNLVKNPDERGFTPLIWASA 133
 Db TTTKGN-----TALHIALAGQDEVVRELNVYGAN-VNAQSQKQGTPLYYAAQ 116
 QY 134 FGEIETVRFLEMGADPHILAKERESALSASTGGYTDIVG----- 174
 Db ENHLEVVKELLENGANGQNVATEDGFTPLAVALQOQHENVVAHLINVTGKVKVLPALHIA 176
 QY 175 -----LLLRVDVINYDMNGG 191
 Db ARNDTTRTAAVLLQNDPNEDVLSKTGFTPLHIAAHYENLNVAQLLNRGASVNFPTQNGI 236
 QY 192 TPLVAVRGVGHVKCVALLARGADLTLEADSGVTPMDLVALGYRKVQVQVIENH 245
 Db TPLHIASRGRNVLMVRLLDLRGAQIETRTKDELTPHCAARNGHVRISILLDH 290
 RESULT 14
 PSDA MOUSE
 ID PSDA MOUSE STANDARD; PRT; 231 AA.
 AC Q92X2; Q9D7N8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 26S proteasome non-ATPase regulatory subunit 10 (26S proteasome
 regulatory subunit p28) (Gankyrin).
 GN PSMD10.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Higashitsuji H., Fujita J.;
 RT "Cloning of mouse gankyrin containing ankyrin repeats";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 29, 2004, 06:16:23 ; Search time 45 Seconds
(without alignments)
1822.994 Million cell updates/sec

Title: US-09-840-243C-11
Perfect score: 1341
Sequence: 1 MELTQPAEDLIQTQTPASE.....VIENHILKLFQSNLVPADPE 260

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	611.5	45.6	296	13 Q7T331	Q7L331 brachydanio
2	439	32.7	144	11 Q8C4M9	Q8C4m9 mus musculus
3	274	20.4	234	5 Q9VL58	Q9VL58 drosophila
4	219.5	16.4	426	16 Q8YT39	Q8YT39 anabaena sp
5	215	16.0	1219	11 Q8C8R3	Q8C8R3 mus musculus
6	214	16.0	737	4 Q8NDR5	Q8NDR5 homo sapien
7	214	16.0	1786	5 Q17344	Q17344 caenorhabdi
8	214	16.0	1809	5 Q17487	Q17487 caenorhabdi
9	214	16.0	1815	5 Q17488	Q17488 caenorhabdi
10	214	16.0	1841	5 Q8MQG0	Q8MQG0 caenorhabdi
11	214	16.0	1867	5 Q17486	Q17486 caenorhabdi
12	214	16.0	2039	5 Q17489	Q17489 caenorhabdi
13	214	16.0	6994	5 Q17343	Q17343 caenorhabdi
14	214	16.0	6994	5 Q17490	Q17490 caenorhabdi
15	213.5	15.9	917	11 Q8HB2	Q8HB2 mus musculus
16	212	15.8	1762	11 Q9EQG6	Q9EQG6 rattus norv

17	212	15.8	1777	4 Q9ULH0	Q9ULH0 homo sapien
18	210.5	15.7	1159	5 Q9NCP8	Q9NCP8 drosophila
19	210.5	15.7	1599	11 Q99NH0	Q99NH0 mus musculus
20	209.5	15.6	2486	4 Q75179	Q75179 homo sapien
21	208	15.5	1693	11 Q8OTG7	Q8OTG7 mus musculus
22	207.5	15.5	1031	4 Q9UF42	Q9UF42 homo sapien
23	207.5	15.5	1715	11 Q9ERD4	Q9ERD4 rattus norv
24	205.5	15.3	321	5 Q9VQI1	Q9VQI1 drosophila
25	205	15.3	448	11 Q8BTQ7	Q8BTQ7 mus musculus
26	205	15.3	469	11 Q8OXC9	Q8OXC9 mus musculus
27	205	15.3	505	11 Q8BXZ6	Q8BXZ6 mus musculus
28	205	15.3	555	11 Q9CZK6	Q9CZK6 mus musculus
29	205	15.3	655	11 Q8OX46	Q8OX46 mus musculus
30	202.5	15.1	765	4 Q8NFD2	Q8NFD2 homo sapien
31	201.5	15.0	1680	13 Q7T163	Q7T163 brachydanio
32	201	15.0	616	4 Q96GK0	Q96GK0 homo sapien
33	199	14.8	454	11 Q8C445	Q8C445 mus musculus
34	199	14.8	1604	5 Q9W2I0	Q9W2I0 drosophila
35	198.5	14.8	1089	5 Q7YU92	Q7YU92 drosophila
36	198.5	14.8	1863	4 Q7Z3L5	Q7Z3L5 homo sapien
37	198	14.8	418	11 Q8BNC1	Q8BNC1 mus musculus
38	198	14.8	627	4 Q8WY90	Q8WY90 homo sapien
39	198	14.8	1486	4 Q8TEF1	Q8TEF1 homo sapien
40	198	14.8	2542	4 Q81WZ3	Q81WZ3 homo sapien
41	198	14.8	2617	4 Q81WZ2	Q81WZ2 homo sapien
42	198	14.8	4001	5 Q8WRQ7	Q8WRQ7 drosophila
43	198	14.8	4001	5 Q9VCA8	Q9VCA8 drosophila
44	197	14.7	686	4 Q9NKF0	Q9NKF0 homo sapien
45	197	14.7	843	11 P97582	P97582 rattus norv

ALIGNMENTS

RESULT 1

Q7T331	PRELIMINARY;	PRT;	296 AA.
ID	Q7T331		
AC	Q7T331;		
DT	01-OCT-2003 (TRENBLrel. 25, Created)		
DT	01-OCT-2003 (TRENBLrel. 25, Last sequence update)		
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)		
DE	Hypothetical protein.		
OS	Brachydanio rerio (Zebrafish) (Danio rerio).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
OC	Cyprinidae; Danio.		
OX	NCBI_TaxID=7955;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Kidney;		
RX	MEDLINE=22388257; PubMed=12477932;		
RA	Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Strausberg R.L., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywinski M.I., Skalska U., Smalhus D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences."		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		


```

RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AR003626; AAF52838.2; -.
DR EMBL; AY061547; AAL29095.1; -.
DR HSSP; P80144; 2MYO.
DR FlyBase; Fgn0032171; CG5846.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 3.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW ANK repeat; Repeat.
SQ
SEQUENCE 234 AA; 24838 MW; A4E9DFBF97BF0514 CRC64;

Query Match
Best Local Similarity 20.4%; Score 274; DB 5; Length 234;
Matches 80; Conservative 37; Mismatches 96; Indels 24; Gaps 6;

QY 13 TQOTPAELGDPEDGEEAAGSDTVVLSLFTCTPEPPVNPPEPDASVSSPQAGSLKHSTT 72
Db 7 TQOTPAELGDPEDGEEAAGSDTVVLSLFTCTPEPPVNPPEPDASVSSPQAGSLKHSTT 72
QY 73 LNRQRGN-EVSALPATLDSLSIHQAQGEHL--DOLKEHLRKGDNLVNKNPDERGFTPLI 129
Db 73 LNRQRGN-EVSALPATLDSLSIHQAQGEHL--DOLKEHLRKGDNLVNKNPDERGFTPLI 129
QY 48 LTNLQRGNTEATFCPEVE-SLSFHERAGQGEITEEQVAAERARQQNIDYK-DAHGFALH 105
Db 48 LTNLQRGNTEATFCPEVE-SLSFHERAGQGEITEEQVAAERARQQNIDYK-DAHGFALH 105
QY 130 WASAFGEITVFLEWGDADPHILAKERESALSLASTGGYTDIVGLLLERDVIDINIYDWN 189
Db 130 WASAFGEITVFLEWGDADPHILAKERESALSLASTGGYTDIVGLLLERDVIDINIYDWN 189
QY 106 WAASVGLVSVQLLVAGANVNTAPDLISPLLLAAGGHNELVFLLEHGADSGHMDIV 165
Db 106 WAASVGLVSVQLLVAGANVNTAPDLISPLLLAAGGHNELVFLLEHGADSGHMDIV 165
QY 190 GTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKVQVQVNIENHI 246
Db 190 GTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKVQVQVNIENHI 246
QY 166 GNTALMYAAGNHPHTCNELLAKDLDSATNEDGDTAYSILAVERGAHLAQAALLEQVM 222
Db 166 GNTALMYAAGNHPHTCNELLAKDLDSATNEDGDTAYSILAVERGAHLAQAALLEQVM 222

RESULT 4
ID Q8YTG9 PRELIMINARY; PRT; 426 AA.
AC Q8YTG9;
DT 01-WAR-2002 (TREMELrel. 20, Created)
DT 01-WAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein A112748.
GN ALL2748.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003590; BAB74447.1; -.
DR FIRM; AE2149; AE2149.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 12.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 13.
DR PROSITE; PS50088; ANK_REPEAT; 11.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
KW Hypothetical protein; Complete proteome.
SQ
SEQUENCE 426 AA; 43800 MW; CF0B9E5D43ACAA03 CRC64;

Query Match
Best Local Similarity 16.4%; Score 219.5; DB 15; Length 426;
Matches 62; Conservative 34; Mismatches 71; Indels 35; Gaps 4;

RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AR003626; AAF52838.2; -.
DR EMBL; AY061547; AAL29095.1; -.
DR HSSP; P80144; 2MYO.
DR FlyBase; Fgn0032171; CG5846.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 3.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW ANK repeat; Repeat.
SQ
SEQUENCE 234 AA; 24838 MW; A4E9DFBF97BF0514 CRC64;

Query Match
Best Local Similarity 20.4%; Score 274; DB 5; Length 234;
Matches 80; Conservative 37; Mismatches 96; Indels 24; Gaps 6;

QY 13 TQOTPAELGDPEDGEEAAGSDTVVLSLFTCTPEPPVNPPEPDASVSSPQAGSLKHSTT 72
Db 7 TQOTPAELGDPEDGEEAAGSDTVVLSLFTCTPEPPVNPPEPDASVSSPQAGSLKHSTT 72
QY 73 LNRQRGN-EVSALPATLDSLSIHQAQGEHL--DOLKEHLRKGDNLVNKNPDERGFTPLI 129
Db 73 LNRQRGN-EVSALPATLDSLSIHQAQGEHL--DOLKEHLRKGDNLVNKNPDERGFTPLI 129
QY 48 LTNLQRGNTEATFCPEVE-SLSFHERAGQGEITEEQVAAERARQQNIDYK-DAHGFALH 105
Db 48 LTNLQRGNTEATFCPEVE-SLSFHERAGQGEITEEQVAAERARQQNIDYK-DAHGFALH 105
QY 130 WASAFGEITVFLEWGDADPHILAKERESALSLASTGGYTDIVGLLLERDVIDINIYDWN 189
Db 130 WASAFGEITVFLEWGDADPHILAKERESALSLASTGGYTDIVGLLLERDVIDINIYDWN 189
QY 106 WAASVGLVSVQLLVAGANVNTAPDLISPLLLAAGGHNELVFLLEHGADSGHMDIV 165
Db 106 WAASVGLVSVQLLVAGANVNTAPDLISPLLLAAGGHNELVFLLEHGADSGHMDIV 165
QY 190 GTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKVQVQVNIENHI 246
Db 190 GTPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMDLAVALGYRKVQVQVNIENHI 246
QY 166 GNTALMYAAGNHPHTCNELLAKDLDSATNEDGDTAYSILAVERGAHLAQAALLEQVM 222
Db 166 GNTALMYAAGNHPHTCNELLAKDLDSATNEDGDTAYSILAVERGAHLAQAALLEQVM 222

RESULT 5
ID Q8C9R3 PRELIMINARY; PRT; 1219 AA.
AC Q8C9R3;
DT 01-WAR-2003 (TREMELrel. 23, Created)
DT 01-WAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Ankyrin 2 (Fragment).
GN ANK2 OR A1835472.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=223545683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK044634; BAC32012.1; -.
DR MGI; MGI:88025; Ank2.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR00906; ZUS.
DR Pfam; PF00023; ank; 24.
DR Pfam; PF00791; ZUS; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 23.
DR SMART; SM00218; ZUS; 1.
DR PROSITE; PS50088; ANK_REPEAT; 20.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
FT NON TER 1219 1219
SQ
SEQUENCE 1219 AA; 131201 MW; AFB0462967AD6184 CRC64;

Query Match
Best Local Similarity 16.0%; Score 215; DB 11; Length 1219;
Matches 55; Conservative 43; Mismatches 79; Indels 20; Gaps 3;

QY 49 PVNPEPDASVSSPQAGSLKHSTTITNQRGNEVSALPATLDSLSIHQAQGEIDOLKE 108
Db 49 PVNPEPDASVSSPQAGSLKHSTTITNQRGNEVSALPATLDSLSIHQAQGEIDOLKE 108
QY 4 PTSFGPEGACTPQNPPIRQSDS-----NASFLAARAGNLDKVV 45
Db 4 PTSFGPEGACTPQNPPIRQSDS-----NASFLAARAGNLDKVV 45
QY 109 HLRKGNLVNKNPDERGFTPLIWAFAFGEITVFLEWGDADPHILAKERESALSASTGG 168
Db 109 HLRKGNLVNKNPDERGFTPLIWAFAFGEITVFLEWGDADPHILAKERESALSASTGG 168
QY 46 YLKGID-INTCNQNLNALHLAKEGVGLVQELLGSGSSVDSATKGNTHIASLAG 104
Db 46 YLKGID-INTCNQNLNALHLAKEGVGLVQELLGSGSSVDSATKGNTHIASLAG 104
QY 169 YTDIVGLLLERDVIDINIYDNGGTPPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMD 228
Db 169 YTDIVGLLLERDVIDINIYDNGGTPPLLYAVRGNHVKVEALLARGADLTTEADSGYTPMD 228
QY 105 QAEVVKVLKVEGANINAGSQNGFTPLYMAQENHIDVVVKYLENGANGSTATEDGTPLA 164
Db 105 QAEVVKVLKVEGANINAGSQNGFTPLYMAQENHIDVVVKYLENGANGSTATEDGTPLA 164
QY 229 LAVALGYRK-VQOVNIEN 244
Db 229 LAVALGYRK-VQOVNIEN 244
QY 165 VALQGHNGQAVAILLEN 181
Db 165 VALQGHNGQAVAILLEN 181

```

```
RESULT 6
Q8NDR5
ID Q8NDR5 PRELIMINARY; PRT; 737 AA.
AC Q8NDR5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFP547039.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Bloecher H., Boscher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL831903; CAD38571.1; -.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0003684; F:damaged DNA binding; IEA.
DR GO; GO:0006298; F:mismatch repair; IEA.
DR InterPro; IPR00432; Muts_C.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 9.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 10.
DR PROSITE; PSS0088; ANK_REPEAT; 9.
DR PROSITE; PSS0297; ANK_REPEAT; 1.
KW Hypothetical protein; ANK repeat; Repeat.
FT NON TER 737 737
SQ SEQUENCE 737 AA; 79652 MW; 75F38DB3AF500175 CRC64;

Query Match 16.0%; Score 214; DB 4; Length 737;
Best Local Similarity 25.8%; Pred. No. 2.3e-09;
Matches 72; Conservative 42; Mismatches 98; Indels 30; Gaps 8;

QY 10 LIQIQQ-----TPASELGDPEDEEADG-----SDTVVLSLFPCTPEPVNPEPDASVS 59
DB 250 LTFQOGLMVASPAQTNDTLDDIMAAVSGRASMSNTPTHTSAASTSQPTPTP-SPII 308
QY 60 SPQAGSSLKHSSTLTNRQRNEVSALPATLDSLSIHQAAQGLDQLKEHLRGKGNLVNK 119
DB 309 SPSAMLLIYPAIDIDAQTESNHDYAL-----TLACAGGHEELVQTLRGASIEH 358
QY 120 PDERGFTPLIWASAFGEIETVRFLEWGADPHILA-KERESALSASTGGYTDIVGLLLE 178
DB 359 RDKKGFTPLIAATAGHVGVEILLDNGADIEAQSERKTDTPLSLACSGRQEVVELLA 418
QY 179 RDVD---INLYDMNGGTPPLIYAVRGNHVKVEALLARGADLTTEADS--GYTPMDLVAL 233
DB 419 RGNKEHNRVSDY---TFLSLAASGGYVNIKIILINAGAEINSGTSGKLGIPLMLAAMN 475
QY 234 GY 235
DB 476 GH 477

RESULT 7
Q17344
ID Q17344 PRELIMINARY; PRT; 1786 AA.
AC Q17344;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE UNC-44 (Fragment).
GN UNC-44.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.

Query Match 16.0%; Score 214; DB 5; Length 1786;
Best Local Similarity 27.3%; Pred. No. 8.7e-09;
Matches 62; Conservative 36; Mismatches 87; Indels 42; Gaps 4;

QY 19 SELGDPEDPCEEAADGSDTVVLSLFPCTPEPVNPEPDASVSSPQAGSSLKHSSTLTNRQR 78
DB 2 SNEGDPDPPQQQQPQESQVQA-----PAAPEPGRAE----- 32
QY 79 GNEVSALPATLDSLSIHQAAQGLDQLKEHLKGNLVNKNPDERGFTPLIWASAFGEIE 138
DB 33 -----GSAFLRAARAGDLEKLELLRAGTD-INTSNANGLNSLHLSKEGHS 80
QY 139 TVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLEDRVDINLYDMNGGTPPLIYAV 198
DB 81 VVRELIKRQAQVDAATRKGNLTALHIALAGQSLIVTLVENGANVNVQSYNGFTPLYMAA 140
QY 199 RGNHVKVEALLARGADLTTEADSGYTPMDLVALGY-RKVQOVIN 244
DB 141 QENHEEVVKLLKHGANQALSTEDGFTPLAVALQQGHDRVVAVLEN 187

RESULT 8
Q17487
ID Q17487 PRELIMINARY; PRT; 1809 AA.
AC Q17487;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE E. elegans ankyrin-related unc-44 (GB:U21734).
GN B0350.2 OR UNC-44.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
```

```
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=95263663; PubMed=7744957;
RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,
RA Boontrakulpoontawe P., Jeyaprakash A., Hedgecock E., Wheaton V.I.,
RA Sobery A.;
RT "An ankyrin-related gene (unc-44) is necessary for proper axonal
RT guidance in Caenorhabditis elegans.";
RL J. Cell Biol. 129:1081-1092(1995).
DR EMBL; U21734; AAA85854.1; -.
DR HSSP; P42773; 11HB.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000488; Death.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00023; ank; 24.
DR Pfam; PF00531; death; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 21.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00218; ZU5; 1.
DR PROSITE; PSS0088; ANK_REPEAT; 22.
DR PROSITE; PSS0297; ANK_REPEAT; 1.
DR PROSITE; PSS0017; DEATH_DOMAIN; 1.
DR PROSITE; PSS0572; GLYCOSYL_HYDROL_F1_1; 2.
KW ANK repeat; Repeat.
FT NON TER 1786 1786
SQ SEQUENCE 1786 AA; 195550 MW; 03E220FB521747F1 CRC64;

Query Match 16.0%; Score 214; DB 5; Length 1786;
Best Local Similarity 27.3%; Pred. No. 8.7e-09;
Matches 62; Conservative 36; Mismatches 87; Indels 42; Gaps 4;

QY 19 SELGDPEDPCEEAADGSDTVVLSLFPCTPEPVNPEPDASVSSPQAGSSLKHSSTLTNRQR 78
DB 2 SNEGDPDPPQQQQPQESQVQA-----PAAPEPGRAE----- 32
QY 79 GNEVSALPATLDSLSIHQAAQGLDQLKEHLKGNLVNKNPDERGFTPLIWASAFGEIE 138
DB 33 -----GSAFLRAARAGDLEKLELLRAGTD-INTSNANGLNSLHLSKEGHS 80
QY 139 TVRFLEWGADPHILAKERESALSASTGGYTDIVGLLLEDRVDINLYDMNGGTPPLIYAV 198
DB 81 VVRELIKRQAQVDAATRKGNLTALHIALAGQSLIVTLVENGANVNVQSYNGFTPLYMAA 140
QY 199 RGNHVKVEALLARGADLTTEADSGYTPMDLVALGY-RKVQOVIN 244
DB 141 QENHEEVVKLLKHGANQALSTEDGFTPLAVALQQGHDRVVAVLEN 187

RESULT 8
Q17487
ID Q17487 PRELIMINARY; PRT; 1809 AA.
AC Q17487;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE E. elegans ankyrin-related unc-44 (GB:U21734).
GN B0350.2 OR UNC-44.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
```


SQ SEQUENCE 1815 AA; 198956 MW; 3ECC6E310915C915 CRC64;
 Query Match 16.0%; Score 214; DB 5; Length 1815;
 Best Local Similarity 27.3%; Pred. No. 8.9e-09;
 Matches 62; Conservative 36; Mismatches 87; Indels 42; Gaps 4;
 QY 19 SELGDPEDGEEAAGSDTVVLSLFPCTPBPVNPEDASVSSPOAGSSLKHSHTLTNRQR 78
 Db 2 SNEGDPFQQOQOQPESEVOA-----PAAPEPGRAE----- 32
 QY 79 GNEVSALPATLDSLSIHQLAAQGLDQLKEHLKGDNLVKNPKDPERGFTPLIWASAFGEIE 138
 Db 33 -----GSASFURAARAGDLEKVLLELRAGTD-INTSNANGLSHLASKEGHS 80
 QY 139 TVRFLEWGADPHILAKERESALSASTGGTYDITVGLLLERDVDTINYYDNGGTPLIYAV 198
 Db 81 VRELKROAQVDAATRKGNLTALHSLAGSLVITLVENGANVNVQSVNGFTPLYMAA 140
 QY 199 RGNHVKVEALLARGADLTTEADSGYTPMDLVALGY-RKVQOVV 244
 Db 141 QENHEEVVKYLLKHGANQALSTEDGFTPLAVALQOQHDRVAVLLEN 187

RESULT 10

Q8MQGO PRELIMINARY; PRT; 1841 AA.
 AC Q8MQGO;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE UNC-44 protein (corresponding sequence B0350.2e).
 GN B0350.2 OR UNC-44.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=9069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Gattung S.;
 RT "The sequence of C. elegans cosmid B0350.";
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR ENBL; U50071; AM75382.1; --
 DR WormRep; B0350.2e; CE31147.
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0007165; P:signal transduction; IEA.
 DR InterPro; IPR002110; ANK
 DR InterPro; IPR000486; Death.
 DR InterPro; IPR001360; Glyco_hydro_1.
 DR InterPro; IPR000906; ZU5.
 DR Pfam; PF00023; ank; 24.
 DR Pfam; PF00791; ZU5; 1.
 DR Pfam; PF00531; death; 1.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SM00248; ANK; 23.
 DR SMART; SM00005; DEATH; 1.
 DR SMART; SM00218; ZU5; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 22.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS50017; DEATH DOMAIN; 1.
 DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 2.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 1841 AA; 201234 MW; 24CC6020DDCE0FC8 CRC64;
 Query Match 16.0%; Score 214; DB 5; Length 1841;
 Best Local Similarity 27.3%; Pred. No. 9.1e-09;
 Matches 62; Conservative 36; Mismatches 87; Indels 42; Gaps 4;
 QY 19 SELGDPEDGEEAAGSDTVVLSLFPCTPBPVNPEDASVSSPOAGSSLKHSHTLTNRQR 78
 Db 2 SNEGDPFQQOQOQPESEVOA-----PAAPEPGRAE----- 32
 QY 79 GNEVSALPATLDSLSIHQLAAQGLDQLKEHLKGDNLVKNPKDPERGFTPLIWASAFGEIE 138
 Db 33 -----GSASFURAARAGDLEKVLLELRAGTD-INTSNANGLSHLASKEGHS 80
 QY 139 TVRFLEWGADPHILAKERESALSASTGGTYDITVGLLLERDVDTINYYDNGGTPLIYAV 198
 Db 81 VRELKROAQVDAATRKGNLTALHSLAGSLVITLVENGANVNVQSVNGFTPLYMAA 140
 QY 199 RGNHVKVEALLARGADLTTEADSGYTPMDLVALGY-RKVQOVV 244
 Db 141 QENHEEVVKYLLKHGANQALSTEDGFTPLAVALQOQHDRVAVLLEN 187

RESULT 11

Q17486 PRELIMINARY; PRT; 1867 AA.
 AC Q17486; O02516;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE C. ELEGANS ankyrin-related UNC-44 (GB:U21734).
 GN B0350.2 OR UNC-44.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906399;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., R.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gattung S.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=N2;
 RX MEDLINE=95263663; PubMed=7744957;
 RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,
 RA Boontrakulpoontawe P., Jeyaprakash A., Hedgecock E., Wheaton V.I.,
 RA Sobery A.;
 RT "An ankyrin-related gene (unc-44) is necessary for proper axonal
 RT guidance in Caenorhabditis elegans.";
 RL J. Cell Biol. 129:1081-1092(1995).

```

RN  [5]
RC  SEQUENCE FROM N.A.
RA  STRAIN=N2;
RA  Otsuka A.J.;
RL  Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN  [6]
RP  SEQUENCE FROM N.A.
RC  STRAIN=N2;
RA  Otsuka A.J.;
RL  Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR  EMBL; U50071; AAA93444.1; -.
DR  EMBL; U39847; AAB41828.1; -.
DR  HSP; P42773; 1.IHB.
DR  WormPep; B0350.2c; CE06704.
DR  GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR  GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR  GO; GO:0007165; P:signal transduction; IEA.
DR  InterPro; IPR002110; ANK.
DR  InterPro; IPR000488; Death.
DR  InterPro; IPR001360; Glyco_hydro_1.
DR  InterPro; IPR000906; ZUS.
DR  Pfam; PF00023; ank; 24.
DR  Pfam; PF00791; ZUS; 1.
DR  PRINTS; PR01415; ANKYRIN.
DR  SMART; SM00248; ANK; 21.
DR  SMART; SM00005; DEATH; 1.
DR  SMART; SM00218; ZUS; 1.
DR  PROSITE; PS50088; ANK_REPEAT; 22.
DR  PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR  PROSITE; PS50017; DEATH_DOMAIN; 1.
DR  PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 2.
KW  ANK repeat; Repeat.
SQ  SEQUENCE 1867 AA; 204253 MW; 4689A2104623B4C6 CRC64;

Query Match 16.0%; Score 214; DB 5; Length 1867;
Best Local Similarity 27.3%; Pred. No. 9.3e-09;
Matches 62; Conservative 36; Mismatches 87; Indels 42; Gaps 4;

QY 19 SELGDPEDPGEAAAGSDTVVLSLFPCTPEPVNPEPDASVSSPQAGSLKHS*TLTNQR 78
DB 2 SNEGDPPQOQQQPESQEQVA---PAAPEGRAE----- 32
QY 79 GNEVSALPATLDSLSIHQLAAQGLDOLKEHLKGDNLVKNKPDGRGFTPLIMASAFGEIE 138
DB 33 -----GSASPLRAARAGDLEKVLLELRAGTD-INTSNANGLNSLHLASKEGHS 80
QY 139 TVRFLEWGADPHILAKERSALSASTGGYTDIVGLLLERDVVDINIDYNGGFTPLIYAV 198
DB 81 VRELKIQQAQVDATRKGNTHALHLAGQSLIVTLVNGANVNVQSVNGFTPLYMAA 140
QY 199 RGNHVKCVEALLARGADLTTEADSGYTPMDLVALGY-RKVQOVVIE 244
DB 141 QENHEEVVKYLLKHGANAQALSTEDGFTPLAVALQQGHDRVAVLLEN 187

RESULT 12
QI7489
ID Q17489 PRELIMINARY; PRT; 2039 AA.
AC Q17489;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE C. elegans ankyrin-related unc-44 (GB:U21734).
GN B0350.2 OR UNC-44.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RX MEDLINE=94150718; PubMed=7906398;

```

```

RA Bonfield J., Buxton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Gattung S.;
RT "The sequence of C. elegans cosmid B0350."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50071; AAA93443.1; -.
DR PIR; T15347; T15347.
DR HSP; P42773; 1.IHB.
DR WormPep; B0350.2a; CE06702.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000488; Death.
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR000906; ZUS.
DR Pfam; PF00023; ank; 24.
DR Pfam; PF00531; death; 1.
DR Pfam; PF00791; ZUS; 1.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 21.
DR SMART; SM00005; DEATH; 1.
DR SMART; SM00218; ZUS; 1.
DR PROSITE; PS50088; ANK_REPEAT; 22.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 2.
KW ANK repeat; Repeat.
SQ SEQUENCE 2039 AA; 222850 MW; 7722C4C2E6AFD68A CRC64;

Query Match 16.0%; Score 214; DB 5; Length 2039;
Best Local Similarity 27.3%; Pred. No. 1.1e-08;
Matches 62; Conservative 36; Mismatches 87; Indels 42; Gaps 4;

QY 19 SELGDPEDPGEAAAGSDTVVLSLFPCTPEPVNPEPDASVSSPQAGSLKHS*TLTNQR 78
DB 2 SNEGDPPQOQQQPESQEQVA---PAAPEGRAE----- 32
QY 79 GNEVSALPATLDSLSIHQLAAQGLDOLKEHLKGDNLVKNKPDGRGFTPLIMASAFGEIE 138
DB 33 -----GSASPLRAARAGDLEKVLLELRAGTD-INTSNANGLNSLHLASKEGHS 80
QY 139 TVRFLEWGADPHILAKERSALSASTGGYTDIVGLLLERDVVDINIDYNGGFTPLIYAV 198
DB 81 VRELKIQQAQVDATRKGNTHALHLAGQSLIVTLVNGANVNVQSVNGFTPLYMAA 140
QY 199 RGNHVKCVEALLARGADLTTEADSGYTPMDLVALGY-RKVQOVVIE 244
DB 141 QENHEEVVKYLLKHGANAQALSTEDGFTPLAVALQQGHDRVAVLLEN 187

RESULT 13
QI7343
ID Q17343 PRELIMINARY; PRT; 6994 AA.
AC Q17343;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)

```


AC Q8BHB2;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE GLP1 homolog.
 GN 9230102N17RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Body, Head, and Heart;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK049454; BAC33756.1; -
 DR EMBL; AK052174; BAC34869.1; -
 DR EMBL; AK082062; BAC38402.1; -
 DR PIR; PT0675; ET0675
 DR MGD; MGI:1924933; 9230102N17RIK.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR007728; Pre-SET.
 DR InterPro; IPR001214; SET.
 DR InterPro; IPR003606; Zn2-binding.
 DR Pfam; PF00023; ank; 7.
 DR Pfam; PF05033; Pre-SET; 1.
 DR Pfam; PF00856; SET; 1.
 DR PRINTS; PRO1415; ANKYRIN.
 DR SMART; SM00248; ANK; 7.
 DR SMART; SM00468; PreSet; 1.
 DR SMART; SM00317; SET; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 5.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50867; PRE_SET; 1.
 DR PROSITE; PS50280; SET; 1.
 SQ SEQUENCE 917 AA; 101017 MW; 3925B0891108F2F8 CRC64;

Query Match 15.9%; Score 213.5; DB 11; Length 917;
 Best Local Similarity 30.6%; Pred. No. 3.6e-09;
 Matches 57; Conservative 31; Mismatches 81; Indels 17; Gaps 3;

QY 62 QAGSSL-----KHSTTLTNQRGNEVSALP-----ATLD-----SLSIHQLAAGGELD 104
 DB 413 QAGANIDTCSEDPRTPLMEAAENNHLDVYLIKAGAQVDPKDAEGSTCLHLAKKGYD 472

QY 105 QKHLKRGDNLVKNPDERGFTPLIWSAFGEIETVRELLWGAADPHILAKERESALSIA 164
 DB 473 VVQYLLSNGQMDVNCQDDGWTPTMIWATEYKHVELVKLLSKGSDINIRDNEENICLHWA 532

QY 165 STGGYTDIVGLLLERDVINIDYDNGSTPLLYAVRGNHVKVEALLARGADLTTEADSGY 224
 DB 533 AFGGCVDAEILLAAKCDLHAVNIHGSGPLHTAARENRYDCVVLFLSRDSVDVTLKKEGE 592

QY 225 TPMDLA 230
 DB 593 TPLQCA 598

Search completed: April 29, 2004, 06:26:46
 Job time : 48 secs